Order Number 9405522

Computational simulation studies of the role of serine 163 in the binding of calcium ions to the low affinity site in subtilisin BPN'

Factor, Alan David, Ph.D.

City University of New York, 1993

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

A

COMPUTATIONAL SIMULATION STUDIES OF THE ROLE OF SERINE 163 IN THE BINDING OF CALCIUM IONS TO THE LOW AFFINITY SITE IN SUBTILISIN BPN'

by

ALAN DAVID FACTOR

A dissertation submitted to the Graduate Faculty in Biomedical Sciences in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York.

1993

This manuscript has been read and accepted by the Graduate Faculty in Biomedical Sciences in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

August 26, 1993
Date

Chair of Examining Committee

August 30, 1993
Date

Executive Officer

Dr. Joseph Goldfarb

Dr. Joseph N. Kushick

Dr. Morris Kraus

Dr. Roman Osman

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Table of Contents

LIST OF TABLES vi

LIST OF FIGURES vii

GLOSSARY x

I. INTRODUCTION AND AIMS 1

A. Mechanism of Ca^{2+} Binding to Subtilisin Site B 1
Hypotheses
Subtilisin Study
Significance

B. Computational Methods for Studying the Biophysics of Proteins 4

C. Representations of Protein Structure and Dynamics 5

II. BACKGROUND AND SIGNIFICANCE 7

A. Ca^{2+} Binding to Proteins 7
Ca^{2+} Binding Motifs in Proteins
Role of Ca^{2+} Binding in Protein Structure and Function

B. Subtilisin BPN' 14
Structure of Ca^{2+} Binding Site B
Hypothesis Regarding the Modulation of Ca^{2+} Binding Affinity
Protein Engineering of Subtilisin

C. Computer Simulations of Macromolecules 24
Molecular Dynamics and Molecular Mechanics
Deterministic and Stochastic Simulations
Empirical Potential Functions
Models for Treating Solvation Effects on Protein Structure and Function
Boundary Methods for Simulating Solvated Proteins

III. METHODS 45

A. Molecular Dynamics Simulations 45

B. Molecular Dynamics Boundary Methods 46
PG3 and Erabutoxin Model Systems

C. Free Energy Perturbation Method 49

D. Representations of Protein Structure and Dynamics 51
Two Dimensional Plots
Linear Distance Plots and Pseudotorsional Plots
Three Dimensional Plots
Distance Matrices, Hydrogen Bond Matrices and Correlated Dynamics Matrices

IV. RESULTS 58

A. Comparison of Boundary Methods in Simulations of PG3 58
Model System
Boundary Methods and Molecular Dynamics Simulations
Conservation of Total Energy and Temperature Equilibration
Peptide Structure
Peptide-Ion Coordination
Solvent Structure
Molecular Dynamics and Peptide Flexibility

B. Effects of SBMD Boundary on Simulations of Solvated Erabutoxin 81
Solvation of Erabutoxin Xray Structure
Equilibration of Protein Simulations
Effect of Stochastic Boundary Conditions
Protein Structure, Protein Dynamics, Solvent Structure

C. Comparison of *In Vacuo* and SBMD Simulations of Subtilisin BPN' 98
Construction of Solvated Xray Ca^{2+} Binding Site
Equilibration of Protein Structures *In Vacuo* and in Solvent
Comparison of Protein Structures *In Vacuo* and in Solvent
Solvent Effects on Protein Structures
Effect of Stochastic Boundary Conditions on Simulations of Subtilisin

D. Mutation of Serine 163 and Effects on Ca^{2+} Binding Site B in Subtilisin 120
Construction and Equilibration of Mutant and Apo Protein Structures
Free energy Perturbation Method, Sudden Deletion Method
Comparison of Native and Mutant Protein Structures
Protein Structure, Ca^{2+} Binding Site, Molecular Dynamics

V. DISCUSSION 153

A. Accuracy of Stochastic Boundary Method in Molecular Dynamics Simulations 153
PG3 and Erabutoxin Studies

B. Importance of Solvent in Molecular Dynamics of Proteins 154
Role of Solvent in Mechanism of Erabutoxin Binding
Comparison of Solvated Subtilisin Simulations

C. Role of Serine 163 in Structure and Dynamics of Ca^{2+} Site B in Subtilisin 155
Comparison of Effects of Serine Mutation in *Apo* and Ca^{2+} Bound Structures
Structure of Ca^{2+} Binding Site
Solvent Accessibility of Ca^{2+} Binding Site
Role of Hydrogen Bonding Interactions in Determining the Structure of the Binding Site
Role of Serine in Affinity of Ca^{2+} to Site B

VI. CONCLUDING REMARKS 157
Ca^{2+} Affinity of the S163A Mutant Subtilisin
Structure of the Ca^{2+} Bound Low Affinity Site B in Subtilisin

VII. REFERENCES 161

List of Tables

Page	Table
18	1. Distances of oxygen ligands from Ca^{2+} in binding sites of xray structure of subtilisin BPN'.
66	2. Alpha carbon rms positional differences from the starting structure for average structures of each 20ps interval of the PG3 simulations.
67	3. Alpha carbon rms positional differences among the last 20ps average structures of the PG3 simulations and the xray structure.
92	4. Alpha carbon rms positional differences among the xray structure, the minimized starting structures and the last 20ps average structures from the erabutoxin simulations.
111	5. Distances of oxygen ligands coordinating the Ca^{2+} in the *in vacuo* and solvated subtilisin simulations and the xray structure.
122	6. List of simulations performed in the Serine 163 mutation study of subtilisin.
128	7. Alpha carbon rms positional differences among the last 20ps average structures from the mutant and apo subtilisin simulations and the starting structure.
146	8. Distances of oxygen ligands from Ca^{2+} in binding sites in the mutant subtilisin simulations and the starting structure.
151	9. Distances between selected atoms in residues S163, K170 and E195 in mutant and native subtilisins.
152	10. Dynamics covariances between selected atoms in residues S163, K170 and E195 in mutant and native subtilisins.

List of Figures

Page	Figure
17	1. Alpha carbon plot of subtilisin BPN'.
20	2. Ca^{2+} binding sites (A and B) in subtilisin BPN'
25	3. Schematic of relationships between experimental, theoretical and modelling studies.
41	4. Schematic of 2-dimensional periodic image boundary method.
63	5. Temperature plots from simulations of PG3.
64	6. Probability distribution function (PDF) of rms of alpha carbons from the last 20ps of the PG3 simulations. The rms difference is calculated from the average structure of the trajectory.
68	7. LDPs of last 20ps average structures from PG3 simulations and xray structure.
69	8. DMs of last 20ps average structures from PG3 simulations and xray structure.
73	9. Peptide carbonyl oxygen-Ca^{2+} distances in final structures from PG3 simulations and xray structure.
74	10. Probability density functions of distances of peptide oxygens coordinating the Ca^{2+} during the last 20ps of the PG3 simulations.
76	11. Radial distribution functions of solvent (average structure) from the Ca^{2+} during the last 20ps of the PG3 simulations.
78	12. Rms positional fluctuations of selected groups of atoms during last 20ps of PG3 simulations.
79	13. Time series of distances of peptide oxygens to the Ca^{2+} during the PG3 simulations.
80	14. Velocity autocorrelation functions and power spectrums of Ca^{2+} from last 20ps of PG3 simulations.
83	15. Total energy plots from erabutoxin simulations.
84	16. Potential energy plots from erabutoxin simulations.
85	17. Temperature plots from erabutoxin simulations.
86	18. Alpha carbon rms positional differences (from the xray structure) during the erabutoxin simulations.

Page Figure

87 19. DM/HBM/LDP plots of last 20ps average structures from erabutoxin simulations and xray structure.

93 20. DLDP/DCAT plots of last 20ps average structures from erabutoxin simulations and xray structure.

94 21. DLDP/DCAT plots of last 20ps average erabutoxin structures from solvated vs. unsolvated simulations.

96 22. B-value plots calculated from positional fluctuations during erabutoxin simulations.

97 23. Radial distribution function of solvent oxygens from origin during last 20ps of solvated erabutoxin simulations.

101 24. Energy plots from *in vacuo* and solvated subtilisin simulations.

102 25. Temperature plots from *in vacuo* and solvated subtilisin simulations.

103 26. Alpha carbon rms positional differences (from the xray structure) from *in vacuo* and solvated subtilisin simulations.

105 27. LDPs of last 20ps average structures from *in vacuo* and solvated subtilisin simulations and xray structure.

106 28. DMs of last 20ps average structures from *in vacuo* and solvated subtilisin simulations and xray structure.

109 29. DDMs of last 20ps average structures from *in vacuo* and solvated subtilisin simulations and xray structure. The difference is calculated from the xray structure.

113 30. Alpha carbon rms positional differences of the Ca^{2+} binding Site B (all alpha carbons within 12.0Å of the Ca^{2+}) of the solvated subtilisin simulation from the xray structure.

114 31. Time series of distances of ligand oxygens to the Ca^{2+} during the solvated subtilisin simulation.

119 32. HBM plots of last 20ps average structure from the solvated subtilisin simulation and the xray structure.

125 33. Temperature plots from mutant and apo subtilisin simulations.

126 34. Potential energy plots from mutant and apo subtilisin simulations.

127 35. Alpha carbon plots of Ca^{2+} bound and apo native and perturbation mutant subtilisin structures.

Page Figure

130 36. LDP plots of last 20ps average mutant and apo subtilisin structures and starting structure.

131 37. DLDP of average structures from perturbation method vs. sudden deletion method.

132 38. DLDP of native subtilisin (starting structure) vs. average Ca^{2+} bound perturbation mutant structure.

133 39. DLDPs of Ca^{2+} bound vs. apo structures from the native and perturbation mutant simulations.

135 40. DM plots of Ca^{2+} bound and apo structures from native subtilisin simulations.

137 41. DM plots of Ca^{2+} bound and apo structures from perturbation mutant subtilisin simulations.

139 42. DDM plots of perturbation mutant Ca^{2+} bound and apo subtilisin structures and apo native subtilisin structure. Differences are calculated from the native Ca^{2+} bound starting structure.

143 43. DM plots of Ca^{2+} binding sites (residues 145 to 210) from mutant and apo native subtilisin simulations and starting structure.

148 44. CDM plots of the last 20ps from the perturbation mutant and native simulations of Ca^{2+} bound and apo subtilisin.

Glossary

This list contains definitions for the common abbreviations used in the text.

Abbreviation	Definition
PG3	cyclo$\left([\text{-L-Pro-Gly}]_3\right)_2$-$Ca^{2+}$
EB	Erabutoxin b
SBT	Subtilisin BPN'
MD	Molecular Dynamics
SBMD	Deformable Stochastic Boundary Conditions
BOX	Periodic Image Boundary Method
HAR	Harmonic Restraint Boundary Method
(D)LDP	(Difference) Linear Distance Plot
(D)CAT	(Difference) Alpha Carbon Pseudotorsional Plot
(D)RMS	(Difference) Root Mean Square Plot
(D)DM	(Difference) Distance Matrix
(D)HBM	(Difference) Hydrogen Bond Matrix
(D)CDM	(Difference) Covariance Dynamics Matrix

I. INTRODUCTION AND AIMS

A. Mechanism of Ca^{2+} Binding to Subtilisin Site B

It is generally true that the activities (e.g. catalysis) and properties (e.g. Ca^{2+} binding) of proteins involve dynamic motions. The role of specific protein elements in the properties it expresses, such as the carbonyl and carboxyl ligands to Ca^{2+} in subtilisin, is usually understood as providing a structural foundation for its function. It is often not well understood what effect specific protein elements, such as hydrogen bonding interactions in Ca^{2+} binding site B of subtilisin, have on the molecular dynamics underlying the properties expressed by proteins.

This thesis explores the structural elements in the binding of Ca^{2+} in one of the sites of subtilisin BPN'. Specifically, the structure and molecular dynamics of the Ca^{2+} binding site B in the native protein and in a mutant are studied via computational simulations of these solvated structures. Determining the role of structural and dynamics interactions between specific elements of the modulatory Ca^{2+} binding site B of subtilisin in the mechanism of Ca^{2+} binding at this site is the main subject of this thesis.

A second aim of this thesis is the development of methods for the analysis of data obtained from computational simulations of macromolecular systems, and their evaluation by use in the present study.

Hypotheses

This thesis explores the hypothesis that the serine residue 163 (S163) in subtilisin BPN' has a specific role in determining the affinity and selectivity of calcium ion binding to the low affinity ion binding site B on this enzyme. S163 is part of a solvent exposed loop of the protein which is adjacent to the two loops that form the low

affinity binding site B (i.e. the loops containing residues 195-197 and 169-175). In the xray structure of subtilisin BPN' the hydroxyl sidechain of S163 shares a hydrogen bond with the sidechains of E195 and K170. The specific hypothesis to be explored is that these interactions stabilize the conformation of the two loops of the binding site, observed in the xray structure, they restrict the flexibility of these loops, and thus reduce the conformational rearrangement that the loop structures undergo in response to Ca^{2+} binding. This hypothesized role of the serine residue (the restriction of loop flexibility) in determining the Ca^{2+} binding site is possibly the mechanism by which the protein achieves selectivity for binding Ca^{2+} over Mg^{2+} [Sussman, 1989]. This mechanism of selectivity for Ca^{2+}, i.e. the achievement of a larger than optimal Ca^{2+}-ligand coordination distances, would hypothetically also result in a greater solvent exposure for the bound ion than the optimal binding conformation.

The expected observations that would support this hypothesis relate to the new orientation of the two protein loops of Ca^{2+} site B (residues 169-174 and 195-200) that should result from the S163A mutation. This mutation is eliminates the hydrogen bond between the side chain of the serine and the residues K170 and E195. This is expected to affect the geometry of the Ca^{2+} ligands and, therefore, its affinity. Also, an observed increase in the flexibility of the two loops of the binding site in the mutated structure would indicate that the serine residue restricts the conformation of this structure.

Subtilisin Study

The hypothesis described above was explored via computational simulations which compared the structure and dynamics of the solvated calcium binding site B in the native protein and a mutant where serine 163 was replaced by an alanine (S163A). The xray structure of the native subtilisin BPN' provided the starting point for a reduced molecular dynamics simulation using deformable stochastic boundary conditions (SBMD). This simulation resulted in an equilibrated system of the protein structure

within a 26 angstrom sphere of Ca^{2+} site B and ca. 3000 solvent molecules. The S163A mutant was constructed from this structure using a perturbation procedure in which the serine was slowly mutated (computationally) into an alanine. The mutated structure was then equilibrated under the same conditions as the native structure. The first part of the study then was the comparison of the structure and molecular dynamics of the native with the mutant protein (with the Ca^{2+} bound).

To explore the structural changes induced by Ca^{2+} binding and the possible role of serine 163 in modulating these changes, a second part of the study was undertaken to compare the native and mutated systems in the absence of Ca^{2+} in site B. This was done by removing the Ca^{2+} from each structure (computationally) and reequilibrating both systems. The equilibrated apo structures were then compared to each other and to the Ca^{2+} bound structures in order to identify changes in the structure and dynamics that could be attributed to the serine residue, thus yielding insight into the validity of the hypothesis.

Significance

The significance of this work stems from the importance of understanding the mechanism of Ca^{2+} binding to proteins, which is a major mechanism for the modulation of the activity of many physiological systems. In the modulatory Ca^{2+} binding site in subtilisin, selectivity for Ca^{2+} is hypothesized to be a property of the conformational flexibility of the binding site. It has been proposed that the site is in a conformation in which the oxygen ligands for Ca^{2+} binding form a structure that is optimal for Ca^{2+} and larger than the optimal structure for binding Mg^{2+} (which has a smaller radius), and that the selectivity for Ca^{2+} results from restricting the flexibility of the binding site in response to cation binding.

Since ion binding to the protein is determined by the relative affinity for the protein versus solvent, the protein may increase selectivity for Ca^{2+} by binding these

ions with less than optimal affinity. Specifically, while the solvation energy of binding of Ca^{2+} and Mg^{2+} is constant, the binding energy of these ions to the protein is dependent upon the protein's conformation. By binding ions in a conformation that is not optimal for Ca^{2+} the protein can achieve selectivity from the fact that the non-optimal conformation may be less favorable for Mg^{2+} than Ca^{2+}. Thus while the binding conformation may bind Ca^{2+} with an affinity of 10kcal/mol less than the energy of the optimal binding conformation, it may affect Mg^{2+} binding even more, reducing its optimal binding energy by, say 50kcal/mol.

B. Computational Methods of Studying the Biophysics of Proteins

The computational methods used in this thesis include molecular mechanics and molecular dynamics simulations which model the dynamic behavior of complex biomolecules. These molecular dynamics simulations of Ca^{2+} binding site B in subtilisin BPN were performed to yield insight into the role of a specific residue in this protein, and the role of the solvent environment, in the mechanism of Ca^{2+} binding to this structure. The effects on Ca^{2+} binding at this site will be deduced from the structural and dynamics effects of mutating serine 163 to alanine observed in the simulations. These simulations provide data at the molecular level on the role that individual protein residues and solvent molecules have in changes in the structural and dynamics properties of a protein resulting from a point mutation.

The solvation model used in this thesis, i.e., the stochastic boundary method, was calibrated against standard methods used in protein simulations. With this model it was feasible to perform simulations on subtilisin in a solvent environment (older solvation methods are not computationally feasible with a protein as large as subtilisin). Also, this solvation method is applicable to a wider range of phenomena (i.e. chemical reactions) than previous methods and thus, as well as making studies of

larger protein structures feasible, it extends the scope of these simulations into other fields of protein chemistry.

C. Representations of Protein Structure and Dynamics

Developments in the field of computer simulations of biomolecular systems have resulted largely from theoretical advancements (e.g. in the accuracy of empirical potential energy functions). Studies using these methods have become more feasible as a result of advances in computational facilities. As this field has progressed, the problem of analysis of the data resulting from computer simulations was exacerbated. On occasion, researchers in this field have referred to the "horrible" amount of data with which they are faced (Dr. Beveridge, personal comm.). Methods of representing both the structural and dynamics data obtained from these studies have become essential.

A new representation of macromolecular structures, the hydrogen bonding matrix [Factor, 1991], is used to analyze the data obtained from some of the studies in this thesis. The hydrogen bonding matrix representation of protein structures determined by xray crystallographic studies or molecular dynamics simulations provides a simple and clear interpretation of the tertiary structural interactions among important structural elements (hydrogen bonding atoms) in a protein. This representation provides a useful tool for relating structure to function and for the analysis of the large amount of data obtained from molecular dynamics simulations. While advances in computer facilities are expected to increase the scope and range of these computational methods, the development of methods of analysis is important for fully interpreting the data. Advances in representing protein structure and dynamics are thus important for progress in this field. These representations of protein structure are equally useful in the analysis of the protein structures obtained from biophysical

methods such as xray crystallography and spectroscopy (e.g. NMR), as well as molecular modelling.

II. BACKGROUND AND SIGNIFICANCE

A. Ca^{2+} Binding to Proteins

Calcium plays a major role in biochemical processes [Blumenthal, 1980; Brostrom, 1978; Carafoli, 1985]. It regulates the activity of many proteins through its role as a second messenger and intracellular levels of Ca^{2+} are highly regulated by mammalian cells. It is essential for the structural integrity and activity of many proteins. Proteins from a wide range of classes bind calcium ions and the mechanisms by which these proteins bind and respond to Ca^{2+}, often integral to their physiological function, is of fundamental importance to a greater understanding of these proteins.

Calcium plays an important role in biological systems and many of its functions are directly related to its interactions with proteins. It is the fifth most abundant element in the human body with 90% residing in bone and teeth. The other 10%, however, circulating as calcium ions, is far more important metabolically and has an essential role in many physiological processes such as muscle contraction, nerve conduction, blood clotting and cell division [Martin, 1973]. These functions involve Ca^{2+} stabilizing, activating or modulating various proteins. The inorganic chemistry of Ca^{2+} underlies all its biophysical activity [Martin, 1973; Einspahr, 1984]. Its concentration in the body, both intracellularly and extracellularly, is highly regulated so that cells may modulate the amount of Ca^{2+} bound to specific proteins. We are attempting to understand how the binding of Ca^{2+} to proteins affect their structural and dynamic properties in order to gain insight into the mechanisms by which Ca^{2+} functions in the larger physiological systems.

Calcium Binding Sites in Proteins

A distinguishing characteristic of Ca^{2+} binding sites in proteins is their binding affinity toward the ion. Sites are characterized as having a low or high affinity with low meaning on the order of micromolar affinity and high meaning on the order of nanomolar affinity. This characteristic has important implications for the role that each type of Ca^{2+} binding site may have in the protein. High affinity binding sites are usually occupied and removal of the Ca^{2+} from these sites is associated with large conformational changes, denaturation, and complete loss of activity. In contrast, removal of Ca^{2+} from low affinity binding sites is usually associated with more subtle conformational changes and often partial activity is retained by the protein [Bajorath, 1989].

Another characteristic of Ca^{2+} binding sites in proteins is their specificity. When high specificity is expressed by low affinity binding sites it is possible for the binding process to transduce a signal (e.g. increase in intracellular Ca^{2+} concentration) into modulation of the activity of the protein. This type of binding site is described as a regulatory site since response to the external signal (Ca^{2+} concentration) can affect, or regulate, the activity of the protein.

The orientation of protein ligands in the proper structural motif of a Ca^{2+} binding site places enormous restrictions on the geometry of the residues containing the ligands as well as the tertiary structure adjacent to these residues in the primary sequence. As a result of the inorganic chemistry underlying Ca^{2+}-ligand interactions, Ca^{2+} binding sites in proteins have ligands forming an approximate octahedral geometry. The preferred number of ligands binding Ca^{2+} in proteins usually range from 6 to 8 where the ligands form an octahedral or pentagonal bipyramidal geometry. In proteins one ligand site is often shared (i.e. bifurcated binding) by the two oxygens from the side-chain carboxyl group of an aspartate or glutamate residue.

The enthalpy of Ca^{2+} binding is very dependent upon the coordination number, geometry, and distances. Quantum mechanical studies of Ca^{2+} binding to carbonyl and carboxyl ligands [Krauss, 1989] have shown that the enthalpy of ligand binding is a result of both the Ca^{2+}-ligand distances and the geometry of the ligands. In xray crystallographic structures of inorganic Ca^{2+} complexes the preferred complexation is a pentagonal bipyramid of oxygen containing ligands at distances of 2.3-2.4Å from the Ca^{2+}. This geometry is also observed in xray crystals of protein structures.

The number and type of ligands contributed to the Ca^{2+} binding site by the protein, as opposed to solvent ligands, is an important characteristic of the binding site. More protein ligands imply greater desolvation of the Ca^{2+} in the process of binding to the protein. Ligands from charged side-chains of aspartate and glutamate residues will have higher enthalpies of binding to Ca^{2+} than uncharged carboxamide ligands (from asparagine or glutamine residue side-chains) or backbone carbonyl ligands. Binding sites with more of these carboxyl ligands will have a higher enthalpy of binding, compensating for the energetic costs of desolvation, compared to binding sites with more solvent ligands. Conversely, a binding site with a smaller number of protein ligands (2 or 3), with several solvent ligands, may suggest lower affinity. In the second case the Ca^{2+} will be only partially desolvated and, from entropic considerations, may possibly have a larger dissociation constant.

In many cases the protein coordinating ligands in Ca^{2+} binding sites are from residues of a short segment of the primary structure. Sequence patterns and structural motifs have been identified for several families of Ca^{2+} binding proteins such as the EF-hand motif in calmodulin, calbindin, parvalbumin and troponin C [Babu, 1987; Herzberg, 1988; Szebenyi, 1987]. Here a single segment of 12 residues adopts a specific conformation that contains the six coordinating ligands of a Ca^{2+} binding site. The motif includes two helices flanking the ends of this segment and adopting an interhelical angle of ca. 100 degrees. Specific residues and the primary sequence are

believed to play an important role in these single segment Ca^{2+} binding sites. It seems that the segment backbone often adopts a strongly restricted conformation enabling several backbone carbonyl groups to bind the Ca^{2+}. Such conformations have been observed to have specific sequence restrictions requiring glycine or proline at certain residues and the motif is observed to have highly conserved residues at ligand positions and other points in the primary sequence.

Another common structural motif is one where the ligands may be from two or three short loops of the protein. This type of Ca^{2+} binding site is highly variable with respect to solvent exposure of the calcium ion. The protein ligands in the solvent exposed Ca^{2+} binding site B in subtilisin, described below, are from two short loops on the surface of the protein.

Role of Calcium Ion Binding in Protein Structure and Function

By definition, occupation of so-called structural binding sites is required for the protein's stability; removal of calcium from these sites results in structural destabilization with loss of activity accompanying loss of structure. As discussed above, Ca^{2+} binding to regulatory (or modulatory) sites also affects the activity and properties of the protein. One pertinent example is the removal of Ca^{2+} from site B in subtilisin which causes increased susceptibility to autolysis and decreased thermal stability. The binding of Ca^{2+} to such a site also affects conformational flexibility, but the relationship of binding to the protein's structure and dynamics is different from that of the structural binding sites. The protein can accommodate conformational changes in these binding sites without a large perturbation in the conformation of other regions of tertiary structure.

While it is easy to understand how specificity can be associated with high affinity binding sites - a rigid binding site in a conformation optimal for Ca^{2+} binding would be relatively unattractive (i.e. express low affinity) for other ions - the basis of

the specificity expressed by low affinity binding sites is less well understood. Low affinity binding sites which also have low specificity are thought to be more flexible than high affinity (structural) binding sites and, due to this intrinsic flexibility, can more easily accommodate other ions, in particular Mg^{2+} which has a very high intracellular concentration. However, the mechanisms by which regulatory sites (i.e. low affinity, high specificity) bind Ca^{2+} must be different and may be related to the mechanisms by which the binding signal modulates the protein's activity.

Binding of the calcium ion directly affects the structure and dynamics of the residues in the binding site. Structural Ca^{2+} binding sites usually provide the optimum geometry for oxygen ligands to bind to calcium ions. The mechanism by which the binding of calcium, at structural sites, increases the stability of the protein is widely believed to be through the addition of ionic, tertiary structural interactions which restrain the flexibility of the protein and thus, via both enthalpic and entropic mechanisms, make it more resistant to thermal, or chemical, denaturation. Ligands in high affinity Ca^{2+} binding sites in proteins are often from residues from remote parts of the primary sequence. One can see how a tightly bound Ca^{2+} in this structure would contribute greatly to the stability of the native conformation by maintaining the tertiary folding and also how the removal of the Ca^{2+} from this structural binding site would destabilize the protein by directly affecting several parts of the primary sequence and indirectly affecting several other regions of the tertiary structure.

A smaller set, compared to inorganic complexes, of xray crystal structures of Ca^{2+} binding proteins have been determined. Binding sites of these proteins are both of low and high affinity and both low and high selectivity. Intracellular Ca^{2+} binding proteins such as calmodulin and calbindin function in environments of low Ca^{2+} concentrations and thus often contain high affinity binding sites (10^{-5}-10^{-8}M) while extracellular Ca^{2+} binding proteins such as subtilisin and trypsin often contain low affinity binding sites (10^{-3}-10^{-4}M).

Xray crystallographic studies of inorganic Ca^{2+} complexes and Ca^{2+} bound protein crystals have shown that the underlying inorganic chemistry of Ca^{2+}, which determines its properties, leads to similar structural interactions in both complexes [Einspahr, 1984]. Ca^{2+} usually forms complexes with 6-8 coordinating ligands with coordination distances ranging from 2.25-2.5Å. The average distance varies directly as a function of coordination number and goes from 2.35Å for a coordination of 6 to 2.45Å for a coordination of 9. Studies comparing xray crystallographic complexes of inorganic Ca^{2+}-water complexes with Mg^{2+}-water complexes have been done to determine if any differences could suggest mechanisms by which protein binding sites may discriminate between these two ions [Einspahr, 1984]. This selectivity is an essential characteristic of modulatory Ca^{2+} binding sites. The conclusions from these studies emphasize the role of the larger ionic radius and preferred coordination number of Ca^{2+} relative to Mg^{2+} and the more restricted distribution of coordination geometries observed for Mg^{2+} vs. Ca^{2+} complexes. A smaller cavity would favor Mg^{2+} over Ca^{2+} binding and high affinity for Mg^{2+} generally demands a more specific (i.e. less flexible) coordination geometry. This analysis was based on inorganic complexes and is believed to be generalizable to protein structures.

High affinity Ca^{2+} binding sites in proteins have short average coordination distances (averaging 2.3-2.4Å). This reflects their lower coordination number (6 or 7) and the increased enthalpy of binding at shorter distances. Low affinity Ca^{2+} binding sites in proteins, however, have larger average coordination distances (2.6-2.7) with similar numbers of ligands coordinating the Ca^{2+} [Einspahr, 1984].

Proteins may prefer a low coordination number in Ca^{2+} binding sites in order to express selectivity for Ca^{2+} over Mg^{2+}. The ionic radius of Ca^{2+} and Mg^{2+} measured from crystal complexes vary directly as a function of coordination number, with Ca^{2+} ranging from 1.0-1.2Å and Mg^{2+} from 0.6-0.9Å over coordination numbers ranging from 6 to 9 (for Ca^{2+}) and 4 to 8 (for Mg^{2+}) [Einspahr, 1984]. A complex with a low Ca^{2+}

coordination number, say 6, may provide a larger relative binding energy of Ca^{2+} over Mg^{2+} even though the absolute binding energy for Ca^{2+} may be greater with a coordination of 7 or 8. It is the relative affinities, free energy differences of binding among the ions, than determine the selectivity of the binding site. The greatest selectivity for Ca^{2+} may be achieved with the lower coordination number.

Ca^{2+} coordination distances vary over a larger range than Mg^{2+} coordination distances. Ca^{2+} has a greater preference for oxygen containing ligands, more easily accommodates larger multidentate anionic ligands and undergoes more rapid substitution [Einspahr, 1984]. All these properties suggest that cavity size and protein flexibility play important roles in determining ion selectivity.

Williams has suggested that the physical chemistry underlying the role of Ca^{2+} in biological systems is its coordination chemistry and its speed of reaction (binding) [Williams, 1986]. He further suggests that while physiological systems (i.e. proteins) differ in the role which Ca^{2+} plays, the underlying mechanisms of action are dependent upon the "mechanical properties" of Ca^{2+}. Ca^{2+} binding sites are not closely associated with active sites in proteins and Williams suggests that it is via conformational changes that Ca^{2+} binding influences protein activity. It is this ability to induce a conformational change that he calls the mechanical property of the calcium ion.

In this thesis the low affinity binding site B in subtilisin is studied. The hypothesis is that, here too, the mechanism of action of the response to Ca^{2+} binding is through conformational changes, but of a very subtle nature. In this site both the structural and dynamics effects of Ca^{2+} binding are believed to be small. The mechanical theory of Williams of the mechanism of action of Ca^{2+} binding in modulating the activity of proteins is consistent with the hypothesis presented in this thesis for the mechanism of action of Ca^{2+} binding to site B in subtilisin. However the importance of the dynamics effects of Ca^{2+} binding upon the activity of this protein have yet to be explored.

B. Subtilisin BPN'

The subtilisins are a family of homologous proteins that are widely used as a model system for the study of protein structure/function relationships. Several excellent reviews on the subtilisin family of proteases, and serine proteases in general, have been published [Fersht, 1985; Kraut, 1977; Blackburn]. Other reviews on the subtilisin family specifically [Ottesen, 1970; Siezen, 1991], as well as articles on their crystallographic structure [Bott, 1988; Drenth, 1972; Wright, 1969] and their mechanism of action [Smith, 1970; Wells, 1986, 1987] provide an excellent foundation for the study of these proteins, which have emerged as a model system for fundamental questions of protein structure. From this vantage of a well known, highly characterized (both structurally and functionally) model system, the subtilisins are also now an important element in the development of new fields such as protein engineering [Carter, 1989; Estell, 1986;Pantoliano, 1988; Wells, 1988a].

Subtilisin was discovered accidentally as the product of a contaminated agar plate of purified ovalbumin. Released into the agar plate, presumably excreted by Bacillus subtilis, the protein enzymatically digested the ovalbumin. The protease was found to be an endopeptidase. It was then purified, sequenced, and its structure determined crystallographically. Studies of its mechanism of action have classified this family of proteins as serine proteases and they represent an example of convergent evolution with the mammalian serine proteases (e.g., trypsin) [Fresht, 1985; Kraut, 1977]. Several strains of bacteria have been found to produce subtilisins and there are several variants of the original protein isolated. Three variants have been studied extensively; two of them, subtilisin BPN' and subtilisin Novo have identical sequences. A third, subtilisin Carlsberg, shows a large degree of homology, with respect to both structure and function, along with a large difference in primary sequence from the

other two. Subtilisin BPN', used in this study, is produced by Bacillus amyloliquefaciens.

The mechanism of action of the subtilisins is believed to be identical to that of the mammalian serine proteases. The subtilisins are relatively small proteins which require Ca^{2+} for activity. They are monomeric, contain about 275 residues and have no cysteine residues and thus no disulfide bridges. The active site is the only part of the tertiary structure of the subtilisins that is homologous to the corresponding structure of trypsin. Both contain a catalytic triad of residues (Asp-His-Ser) whose orientation enables a proton transfer from Ser to His, ionizing the serine, and resulting in a nucleophilic attack on the backbone carbonyl carbon of the scissile peptide bond of the substrate. A tetrahedral intermediate state (on the substrate carbonyl carbon) forms and, after cleavage of the peptide bond, results in an acyl-enzyme intermediate and an amide leaving group. This mechanism, including the oxyanion hydrogen bond stabilization of the backbone amide nitrogen adjacent to the carbonyl carbon of the substrate, is common to both the bacterial and mammalian families of proteins. Nevertheless, there is very little primary sequence homology between the two families, and even the order of the critical triad of residues involved in the catalytic activity (Asp-His-Ser) in the primary sequence of the two families is different. This strongly supports the concept of two different paths of protein evolution having converged upon one catalytic mechanism [Fersht, 1985].

Subtilisin has several advantages as a subject for the theoretical study of Ca^{2+} binding mechanisms in proteins. First and foremost it is a well defined system. Its native conformation has been determined xray crystallographically at high atomic resolution. Both Ca^{2+} binding sites are structurally well defined. Its activity has been studied thoroughly and its catalytic mechanism of action is well understood. Since it is a serine protease, there are many proteins of various degrees of homology which are available for further study and clarification of hypotheses generated by work on

subtilisin. Theories relating structural and dynamic properties of subtilisin, such as the role of protein flexibility in Ca^{2+} binding, may be tested on trypsin, proteinase K, and the other homologous proteins.

There is a large body of data on the biophysical properties of both native and mutant subtilisins. There is also a large body of data on the response of subtilisin to chemical modifications. These studies have elucidated the role of residues in the active site (the catalytic triad of Asp, His, Ser) as well as other structural features of the protein (e.g. the oxyanion hole). These chemical modification studies, as well as the mutation studies, have provided the foundation for subtilisin to become an important subject of studies in protein engineering. The role of subtilisin as a subject for protein engineering studies provides another reason for choosing this protein as the subject of theoretical studies on protein structure/function relationships as the theoretical work, such as this thesis, can be tested by studies of protein engineered mutants.

Structure of Calcium Binding Site B

Subtilisin BPN' is an A/B (alpha/beta) protein [Richardson, 1981]. It is globular, monomeric, contains 275 residues and no cysteines. There are two Ca^{2+} binding sites - one high affinity (site A) and one low affinity (site B); the calcium ions are over 30Å apart in the xray structure. Figure 1 is an alpha carbon plot of subtilisin BPN'. In site B the Ca^{2+} is highly solvent exposed while in site A it is sequestered from solvent. This is shown by the types of ligands observed in the xray structure (shown in Table 1). Both sites are near the surface of the protein however. The structure is stabilized by an alpha helix that runs through its center. Except for two short anti-parallel beta strands separated by a turn all the beta strand interactions are parallel. The catalytic triad of residues serine 221, histidine 64 and aspartate 32 form an active site that is near one end of the central helix. This active site is next to the substrate binding site which is the A/B/A/B sequence of secondary structures formed by residues 100-165.

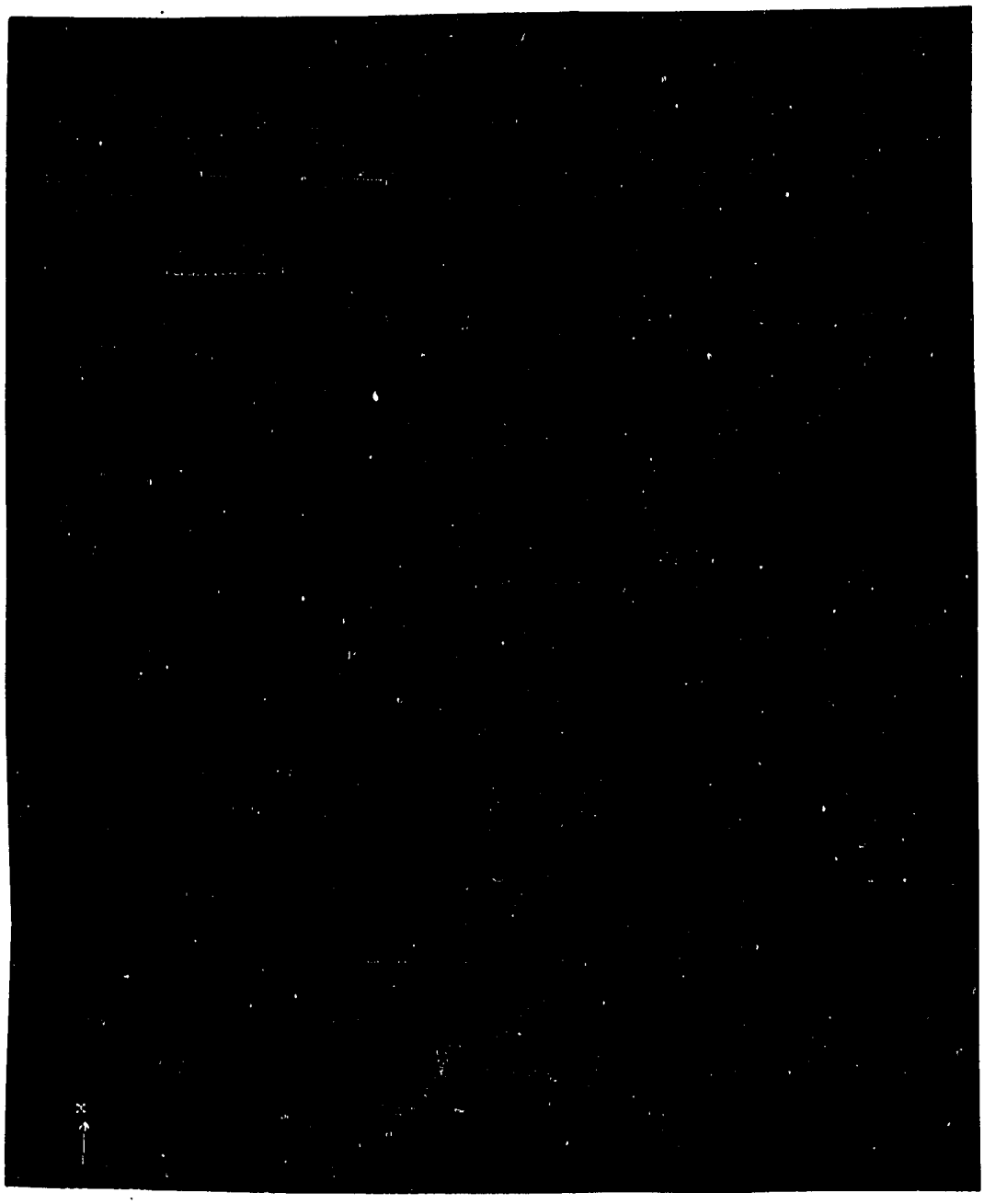


Table 1. Distances of Ca^{2+} ligands in xray structure of subtilisin BPN'.

Distances of Oxygens Coordinated at High Affinity Binding Site A

Residue	Ligand	Distance
GLN2	OE1	2.37
ASP41	OD1	2.42
ASP41	OD2	2.52
LEU75	O	2.29
ASN77	OD1	2.40
ILE79	O	2.29
VAL81	O	2.37

Distances of Oxygens Coordinated at Low Affinity Binding Site B

Residue	Ligand	Distance
GLY169	O	2.66
TYR171	O	2.97
VAL174	O	2.66
GLU195	O	3.02
ASP197	OD1	2.84
WAT42	OH2	2.72
WAT74	OH2	2.97

The high affinity Ca^{2+} binding site is structural; its protein ligands are from the N terminal residue 2, residue 41 and the loop 75-81. This site is approximately on the other side of the active site across from the substrate binding site. The low affinity Ca^{2+} binding site B is much closer to the substrate binding site and adjacent to the loop containing residues 160 to 169. Its protein ligands are from the two loops of residues 169-175 and 195-200.

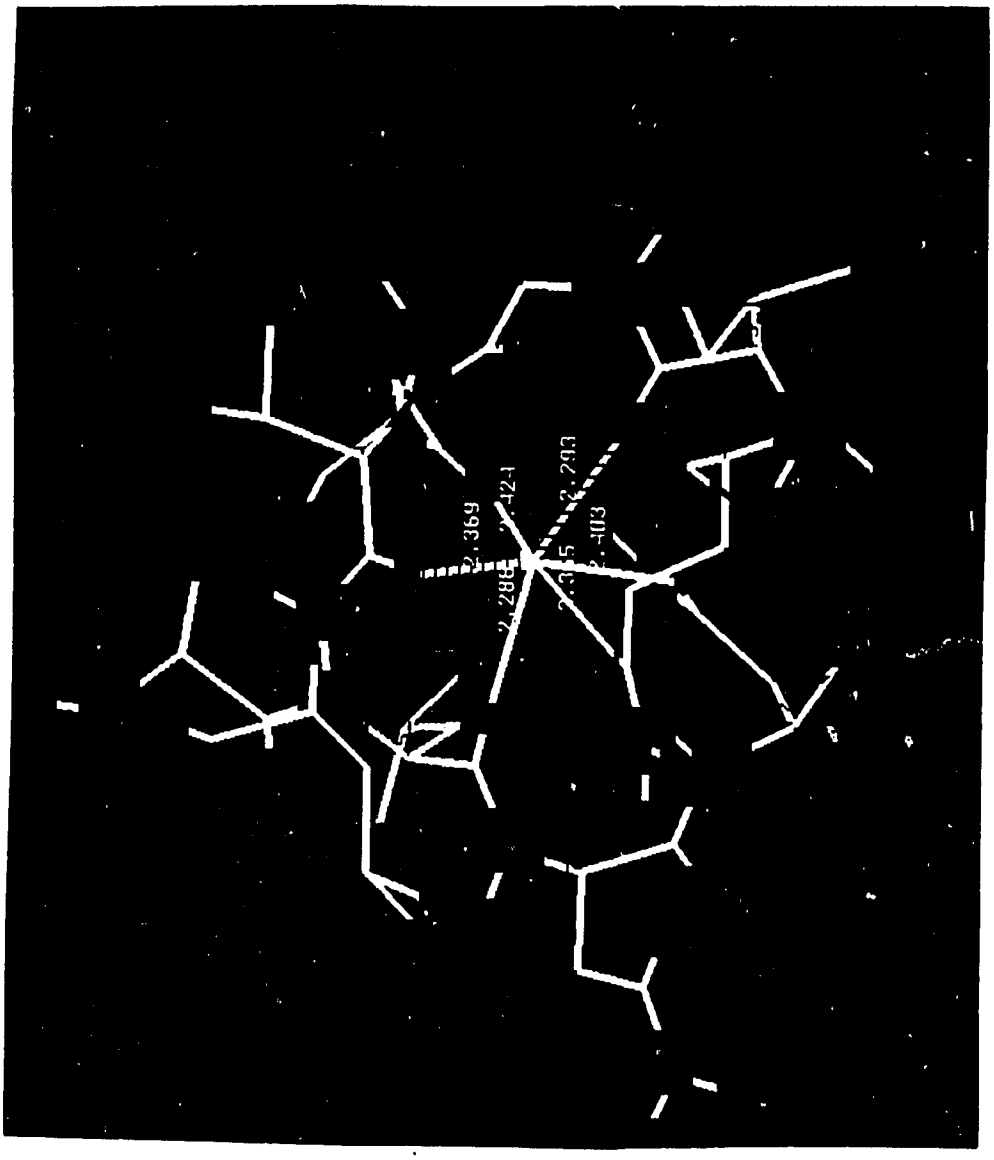
The functional relationship between Ca^{2+} binding site B and the rest of the protein, specifically the active site and the substrate binding site, is not well understood. The site is structurally well defined, consisting of the two protein loops mentioned above, and interacts with adjacent protein residues as well as the Ca^{2+} ion and solvent.

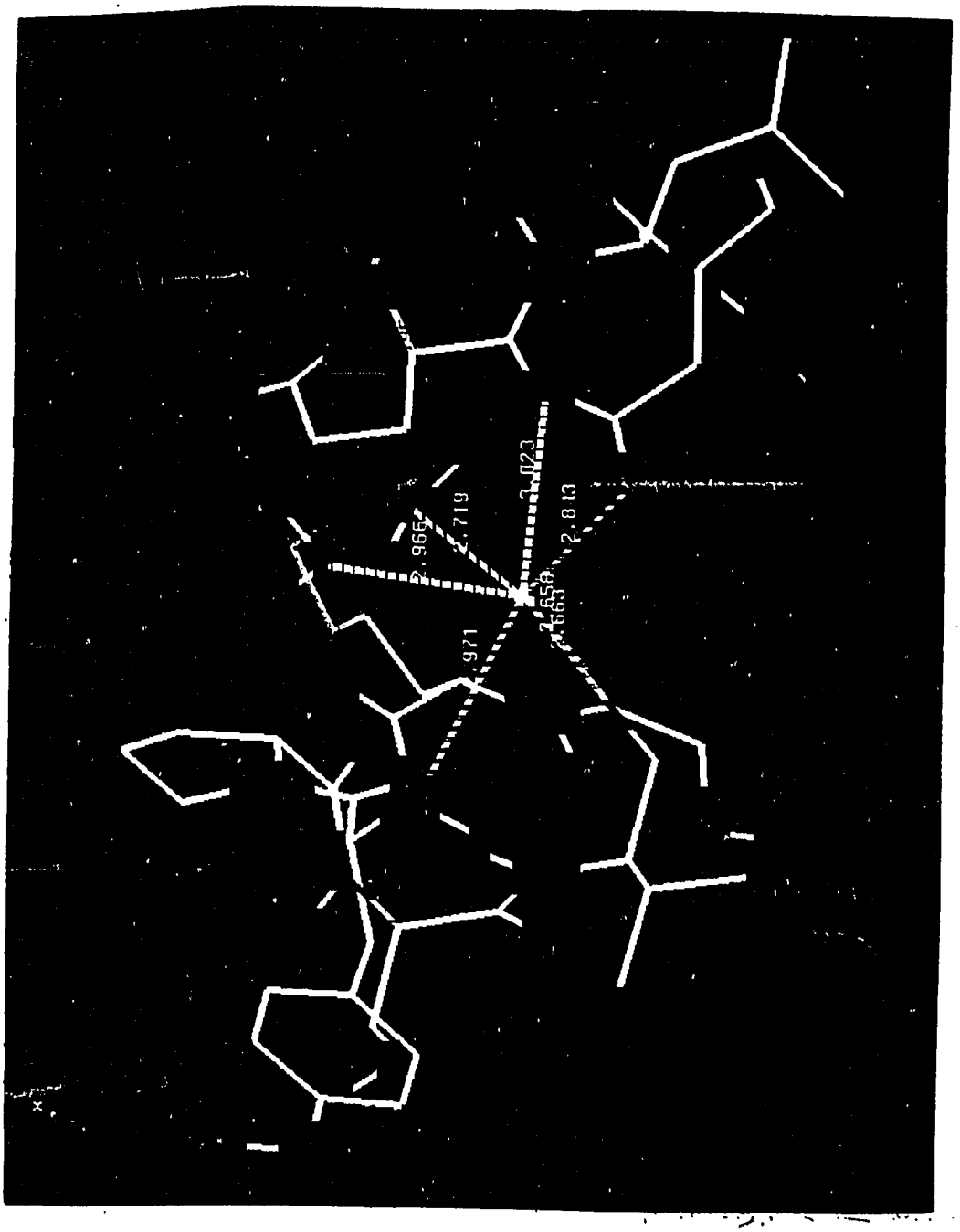
In the xray structure the Ca^{2+} ligands in site B form the usual octagonal coordination geometry. The coordination distances are unusually large, however, even for a protein Ca^{2+} binding site. The distances found in the xray coordination of the high affinity Ca^{2+} site A are significantly shorter than those in site B. All the coordinating ligands contain oxygen and are either carboxyl or carbonyl groups (or solvent). The ligands and distances are shown in Table 1. Figures 2a-b show the Ca^{2+} binding sites. Removal of the Ca^{2+} from the structural binding site A results in complete loss of structure and activity. The effects of removal of the Ca^{2+} from site B are subtle. These are observed to be a decrease in stability against both thermal and chemical denaturants and a slow loss of activity (over a period of hours) due to autocatalysis.

Hypothesis Regarding the Modulation of Calcium Ion Binding Affinity

The structural change resulting from removal of Ca^{2+} from site B is proposed to be caused by greater flexibility resulting from the loss of stabilizing interactions between the Ca^{2+} and the coordinating ligands along with a structural reorientation that is a response to the large repulsion between coordinating ligands in the apo site. Solvent will form stabilizing interactions among protein residues in the binding site, screening the repulsive interactions between the anionic ligands in the Ca^{2+} free structure. The apo structure is hypothesized to be more flexible than the Ca^{2+} bound conformation.

We hypothesize that the conformational change in the rest of the protein, resulting from the removal of the Ca^{2+} from site B, is much smaller than the change in the binding site itself. There are no rigid structural elements such as disulfide bonds or beta sheet interactions connecting the binding site to other regions of the protein so large structural perturbations resulting from the removal of Ca^{2+} will most likely be limited to the binding site. This agrees with the small structural perturbation observed


2.369
2.288
2.293
2.403


3.023
2.966
2.719
2.813
2.971

experimentally as a result of low (10^{-6} M) Ca^{2+} concentrations. Xray crystallographic studies of an homologous protein, proteinase K, with and without Ca^{2+} bound in the low affinity Ca^{2+} binding site of this protein, show that when Ca^{2+} is absent from the low affinity binding site subtle structural changes occur which propagate slowly along a series of adjacent surface loops from the low affinity Ca^{2+} binding site to the substrate binding site [Bajorath, 1985]. We hypothesize that a similar effect (i.e., structural perturbation) may occur in subtilisin after removal of Ca^{2+} from its low affinity Ca^{2+} binding site.

The effects of the loss of Ca^{2+} on the dynamics of the structural elements in the protein adjacent to the binding site are also hypothesized to be an increase in flexibility. This is consistent with an increase in entropy and a decrease in the free energy difference from the denatured state and would agree with the observed decrease in stability in the apo structure. Since these effects are observed experimentally over a time scale of several hours it is improbable that they would be observed in simulation studies. However perturbations in the dynamics of the protein obtained from the simulations, especially dynamical interactions involving the Ca^{2+} binding site, may suggest a basis for the experimentally observed decrease in stability.

The hydrogen bonding interactions between the two protein loops in Ca^{2+} binding site B is proposed to play a key role in maintaining the conformation of the binding site. It has been hypothesized [Krauss, 1989] that the binding site in the xray structure is maintained in a conformation that results in larger than optimal ion-ligand coordination distances as a mechanism for selectively binding Ca^{2+} over Mg^{2+}. The reduction in affinity (binding free energy relative to the solvent) for Ca^{2+}, resulting from the binding site being restricted to the conformation observed in the xray structure, is hypothesized to be much less than the reduction in affinity for Mg^{2+}. This would result in an increase in the selectivity for (relative affinity of) Ca^{2+} over Mg^{2+} at this site. The hydrogen bonding interactions between lysine 170 and glutamate 195, of

the loops in the binding site, with serine 163 are hypothesized to provide the mechanism for stabilizing the unusually large coordination distances observed in the xray structure. Without this stabilization the structure is hypothesized to collapse into a conformation in which the optimal, smaller, coordination shell for Ca^{2+} would be observed.

Protein Engineering of Subtilisin

Advances in DNA technology, that occurred ca. the 1970's, were very important to the development of the field of protein engineering [Leatherbarrow, 1986]. Site-directed mutagenesis is replacing the older methods of chemical and enzymatic modification of proteins as the major method for studying protein structure/function relationships. The ability to produce any mutation in native proteins is far more powerful than the limited modifications possible with the older methods. This holds great possibilities for producing proteins tailored to specific pharmaceutical and chemical applications. However, greater understanding of protein structure/function relationships is required to utilize the power of protein engineering fully. There is an obvious need for work in this field to be guided by the rationale which comes from understanding the fundamental relationship between a protein's structure and the properties and activities it expresses. Indeed without this understanding of the biophysical basis of protein function it is not possible to take full advantage of the opportunities that develop from advances in fields such as protein engineering (e.g., mutagenesis). Recognition of this need for mechanistic understanding should increase demand for studies of protein structure/activity relationships at the molecular level.

Much of the work that has been done in protein engineering involves readily measurable properties. Thus the effect of mutations on protein stability and catalytic activity is often the subject of work in this field. Structure/activity studies of the affinity and selectivity of ligand binding sites is another property often investigated.

Subtilisin has been the subject of numerous structure/function studies involving engineered mutants. Some of the properties studied include oxidative stability [Wells, 1985], substrate specificity [Wells, 1985], and catalytic activity [Carter, 1988]. However this thesis only concerns work relating to the Ca^{2+} binding properties of subtilisin. Studies have shown that many site-directed mutations result in only small structural perturbations localized to the region surrounding the mutation, and that the results of such "structurally conservative" mutations can often be qualitatively, if not quantitatively, predicted. This knowledge is useful as interpretation of results from studies of protein engineered mutants often rely on the assumption that there was a structurally conservative response to the mutation.

C. Computer Simulations of Macromolecules

Computer simulations of macromolecular systems are a form of modelling. As shown in Figure 3 (adapted from Casta, 1989), modelling has fundamental relationships with experimental and theoretical studies. It is important to understand the unique role that modelling has with respect to experiment and theory. In biophysics, computational studies begin with a well-defined physical model. From these simulations both our understanding of the natural physical systems, as well as the accuracy of analytical theories, can be explored.

The fundamental purpose of constructing models of natural systems is to "bring a measure of order to our experience and observation" from which predictions can be made [Casti, 1989]. The quality (and accuracy) of the predictions which can result from a given model are one measure of its value. From this perspective, modelling serves to integrate experimental observations (much as theory does). From the perspective of theory, however, modelling serves as an empirical test ("experiment") when experimental studies are not feasible. Thus models allow the comparison of theory with experiment when direct comparison is not feasible. For example, the microscopic

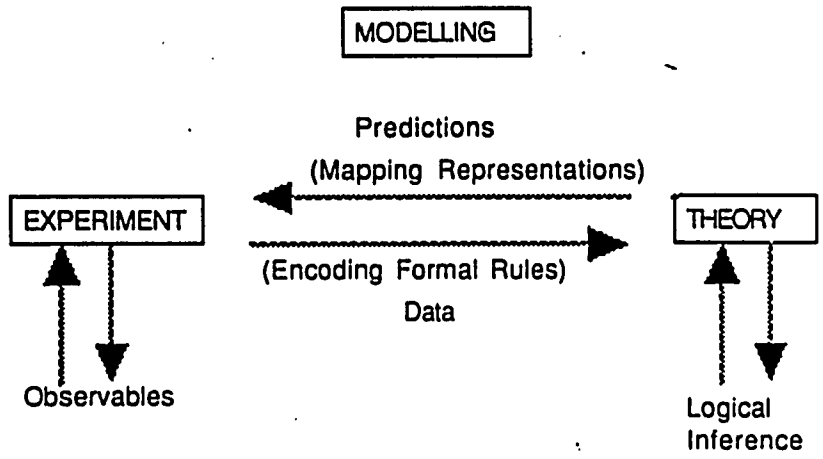


Figure 3. Schematic representation of the relationships between experiment, theory and modelling. Experiments on natural systems provide the data used in constructing theories of these systems. Theory is developed based on the data from exerimental studies and logical inferences from the formal rules that represent the system. Predictions from the theory are tested on representative models of the natural system developed from the theory as well as on the natural system itself.

trajectory, even at the thermodynamic limit, of an oxygen molecule binding to myoglobin, is beyond the most sensitive spectroscopic (experimental) methods. New insights into the mechanism of this reaction has been gained from studies of computational simulations of this process [Karplus, 1986; Kottalam, 1988].

Simulation studies are now performed on a wide range of systems (e.g. proteins, inorganic polymers, and various other chemical systems). The first major computational study of a protein was performed on a small peptide inhibitor of trypsin (BPTI), by Karplus at Harvard [McCammon, 1977], less than twenty years ago; now new technical developments (i.e. methodology) in this field are being done using the protein (trypsin) of which BPTI was only a ligand.

A growing concern has become the inappropriate application of generally available, and easily used, computational chemistry packages without an appreciation of the underlying biophysics or without the careful attention required to the parameters and other variables involved in performing a simulation.

A basic assumption underlying this thesis is the intrinsic value of computational simulations, based on well-defined physical models, to provide data which result in new insights into biophysical processes unattainable by any known experimental method. There are several caveats to this approach which require discussion. First there are the inherent dangers of applying an inaccurate model to a natural system or drawing inappropriate conclusions (overreaching) by going beyond the valid domain of the model. Much of the research that is performed after the introduction of a new method (or the initial application of an existing method to a new natural system) has the goal of insuring the validity of the new application through repeated checks to experimental data. A good example is the application of molecular dynamics to proteins. After many studies, in which this method was applied to simple, homogeneous systems of inert atoms (such as argon), resulted in a large body of data accurately reproducing experimental observations, the validity of these methods was

established for those molecular classes on which this early research was performed. The methods were developed and tested on a limited range of chemical systems and there was no assumption that they could be applied successfully to more complex systems. For the initial molecular dynamics studies on proteins it was necessary, not only to develop a modified empirical potential energy function for that class of molecules, but, also, to establish of the validity of applying these methods to this class of complex macromolecular structures.

Molecular dynamics, and molecular mechanics, are the main computational methods used here [Berendsen, 1991; Cornell, 1991]. These methods are based on classical mechanics. Since the studies presented here do not involve chemical reactions, to a first approximation quantum mechanical effects can be neglected. Both molecular mechanics and molecular dynamics are computational simulations of the deterministic type. In molecular dynamics, for example, given initial values for the configuration of a system, a simulation is performed by computing the atomic trajectory in discrete steps according to classical equations of motion and using an empirical potential function. The equations of motion can be formulated in Newtonian, Hamiltonian or Lagrangian form. In addition to these deterministic methods, a stochastic method was also used to model the solvent environment.

Molecular Dynamics and Molecular Mechanics

Molecular mechanics is a common method of studying the relationship between conformation and energy and optimizing molecular structures [Mackay, 1989]. Excluding external fields, such as boundary constraints, the computed energy of a molecule is determined by the potential function (see Empirical Potential Functions below).

The simplest computational study of a molecular structure is a molecular mechanical analysis in which the energy is calculated for various conformations. A

conformation is the 3-dimensional structure of the molecule defined by the coordinates of all the atoms and their topology, where the topology is the chemical structure (i.e. bonds) of the molecule. Conformations may be produced randomly, which is done in stochastic computational methods, or systematically. The study usually starts from a structure obtained experimentally, mostly from xray crystallography but with increasing contributions from NMR spectroscopy. The starting structure for the first study presented here, on PG3 (see below), was obtained from xray crystallography and NMR. The structures of the proteins used in the other studies, erabutoxin b and subtilisin BPN', were obtained by xray crystallography.

Molecular mechanics is used to study structures obtained from experiment and also to optimize these structures for molecular dynamics simulations. Apart from the determination of the energy of a given conformation, which is common to molecular mechanics and molecular dynamics algorithms, the term molecular mechanics is also applied often to mean an optimization procedure. In optimization procedures various algorithms are used to minimize the total energy of the molecule by iteratively modifying the conformation. These procedures basically calculate the potential energy gradient (forces) on each atom in the molecule (i.e. a molecular mechanics calculation), move the atoms down the gradient (to a lower energy conformation) and repeat this for a given number of steps until the procedure converges or until a tolerance is reached. Usually, an end point for optimization is either reaching a limit for the gradient on the atoms or reaching a limit on the change in energy after a step in the procedure.

Various minimization algorithms are commonly used but none are able to aim directly for the conformation with the global energy minimum. All available minimization algorithms converge into local minima, with the exception of the exhaustive search of all possible conformations (which is not feasible for macromolecules). Optimization procedures can be very sensitive to the starting conformation (i.e. not robust). Fully optimizing any given structure will not overcome

potential energy barriers (as the conformational energy is minimized) and thus cannot efficiently search the conformational space of the structure.

Molecular mechanics is not used to explore the conformational space of proteins since molecular dynamics provides a better method to do this as well as yielding data on the time development of conformational changes. Molecular mechanics was used to partially optimize the structures obtained from experimental methods. This is necessary since the data obtained by xray crystallography is an average over time and over the conformations of all the molecules in the crystal. Added to this source of error is the experimental error introduced by the procedure of determining the structure from the experimental data (e.g. reflections collected from an xray crystal). When an empirical potential energy function is applied to the structures determined from these experimental methods, there are usually large residual forces on individual atoms. This is called strain energy and is removed by optimization of the structure. The conformation obtained after this procedure is usually very similar to the experimentally determined structure.

Molecular dynamics is a natural development of molecular mechanics [Alder, 1959; Karplus, 1990]. Here too the potential energy gradient is calculated but kinetic energy is added to the calculation of the forces on each atom. The time development of the molecule, its trajectory, is simulated with a discrete, iterative algorithm. Generally a molecular dynamics simulation begins with a period of heating in which random velocities (e.g. a gaussian distribution) are added to the atoms. The velocities are scaled periodically to gently increase the temperature from 0°K to 300°K. This usually requires calculating a trajectory of several thousand steps. A discrete timestep is used; the potential gradient on each atom is calculated, the atoms are projected to their new positions assuming a constant velocity over the timestep interval, and the procedure is repeated (i.e. the atoms jump from one position the next). This describes, essentially, the Verlet algorithm [Verlet, 1967] for molecular dynamics simulations. After heating,

the simulation enters an equilibration phase, during which time the temperature of the system is periodically checked and readjusted to the set temperature (e.g. 300°K) by changing the kinetic energy of the system (e.g. scaling the atomic velocities). After the system has equilibrated (i.e. the temperature is stable) the production phase begins and data on the molecular dynamics of the system can be collected.

Molecular dynamics simulations compute the phase space trajectory of a system using the laws of classical mechanics. The complexity of the equations of motion for large systems precludes analytical solutions, so that molecular dynamics uses a numerical integration method and iterative calculations at discrete time intervals to simulate the system's trajectory. As a concrete example, describing the Verlet molecular dynamics algorithm, the following description of simulating a single particle, one-dimensional harmonic oscillator will be given (following Heermann, 1990). While this example is trivial to solve analytically it also yields a clear understanding of the method.

In this example the atom will follow a trajectory forming a surface of constant energy defined by the Hamiltonian describing the force of a spring, with a force constant k, i.e.

$$H = \frac{p^2}{2m} + \frac{1}{2}kx^2 \qquad (1)$$

where p is the momentum, m is the mass, and x is the position of the atom. To apply this (deterministic) method one must start with an initial configuration; here the initial position and momentum of the atom must be given. A timestep, h, is defined, and recursion formulas for the position and momentum of the system are used to compute the motion of the atom. The recursion formulas used in the Verlet algorithm are

$$x(t+h) = 2x(t) - x(t-h) + \frac{F(t)h^2}{m} \qquad (2)$$

$$v(t) = \frac{(x(t+h) - x(t-h))}{m} \qquad (3)$$

where x(t) is the position, v(t) is the velocity, and F(t) is the force on the atom at time t.

The total time (trajectory) that can be feasibly simulated is limited by size of the timestep. The size of the timestep used in the simulation is limited by the physics of the system; it must be small enough for the computed trajectory to accurately model the dynamics of the system. This limitation is related to the fastest motion (i.e. vibrational frequency) of the system. In proteins this is the bond stretching vibrations of polar hydrogens to heavy atoms; it limits the upper bound of the time step to 0.5 femtoseconds (fs). Polar hydrogen bond vibrations are often suppressed in the simulations of proteins as this allows the timestep to be increased (to 1.0fs) and these motions are not significant to the molecular dynamics of these systems. A large drift in total energy (greater than 1%) results from an inaccurately simulated system where the error comes from the timestep being too large. The timestep is only one of many sources of error. Due to the discretization of the calculation (using a finite timestep) the computed trajectory will always differ from the actual trajectory.

The Verlet algorithm, described above, simulates a microcanonical system where the variables N, number of atoms, V, volume, and E, total energy can be considered constant. Different types of behavior may be modeled by the addition of different forces. Another dynamics algorithm (Langevin) used in this thesis models Brownian dynamics where the constants are N, V, and T, temperature. The Langevin algorithm differs from the Verlet algorithm by the addition of two forces which model Brownian motion; a random force whose ensemble average is zero and a frictional force proportional to a given atom's velocity. The frictional force coefficients used in the model are empirical parameters developed from experimental data on the biophysical properties (e.g. viscosity) of the system under study.

Deterministic and Stochastic Simulations

There are three points to consider when proposing a molecular modeling study. One is the accuracy of the empirical potential energy function, and is discussed in a separate section. The other two are extensive properties of the simulation: the phase space sampled and the time scale of the simulation. These two are interrelated. Systems which are the subject of molecular dynamics simulations, such as protein macromolecules, do not yield to analytical treatment of their dynamics. These simulations suffer from limitations on all three fronts, but, of course as the complexity of the system increases so may the value of the information the model yields.

The two extensive limitations on molecular dynamics simulations, the amount of phase space sampled and the time scale of the simulation, are fundamental. While there are methods to extend the range of both the phase space (high temperature dynamics, simulated annealing) and time scale (Brownian dynamics, Monte Carlo simulations), these entail a loss of data from the dynamics results of the simulation. There is a tradeoff in the increased range of the phase space sampled, for instance, with a reduction in information on the molecular dynamics of the system. These two limitations illustrate the differences between stochastic and deterministic methods of modelling macromolecular systems. At the heart of this difference lies the relationship of molecular dynamics to the questions being explored.

In deterministic methods, such as molecular dynamics simulations, the dynamics of the system are the heart of the simulation. The data collected (the trajectory) result from the application of the model (potential functions and equations of motion). In this model, an initial value problem according to classical mechanics, the initial state of the system will determine the results. The main advantage of the deterministic methods over stochastic approaches is that they result in the most accurate simulation of the dynamics of the system - its motion through time.

The stochastic methods are based on a different approach - concepts from probability theory and statistical mechanics. The goal of these methods is the computation of expectation values (i.e. average values) via ensemble averaging over all possible configurations (i.e. conformations) of the system. These methods use probabilistic (e,g. Markov) processes to produce transitions between conformations and have the advantage of being applicable even when dynamics is not an intrinsic element of the model.

In molecular dynamics simulations ergodicity is invoked which allows conclusions about the equilibrium state of a system to be drawn from a finite set of conformations. This means that the ensemble average of a property calculated from a finite trajectory is assumed to be equal to the time average of all conformations (if the system simulated is at equilibrium).

Statistical mechanics postulates that, over time, all of the phase space accessible to the system will be sampled with a distribution that is a function of the energy. The integral of this distribution is a fundamental concept, called the partition function, Z, defined as

$$Z = \int f(H(x)) \tag{4}$$

where the integration is over all of phase space (all conformations of the system). H(x) is the Hamiltonian (energy) of the system in a given conformation, x, and f() is the distribution function of this Hamiltonian (i.e. a function of the probability the system will occupy this conformation). By definition the ensemble average is

$$\langle A \rangle = \frac{\int A(x) f(H(x))}{\int f(H(x))} \tag{5}$$

where, again, the integration is over all of phase space. In deterministic methods, the time average, calculated from the set of conformations obtained from a trajectory, is defined as

$$\overline{A} = \frac{1}{n}\sum_{i=t}^{n} A(x(t)) \qquad (6)$$

where x(t) is the conformation at time t and n is the number of conformations in the trajectory.

Stochastic methods cannot sample all of phase space. Analogously, deterministic methods cannot produce an infinite time set. Limited sets of conformations are produced by both methods

When modelling molecular dynamics an essential question is whether the trajectory that is obtained has adequately explored the system in the area of interest (e.g. local conformational space). This question is important from a time perspective as well as a spatial perspective. Simulations of large biomolecular systems are currently limited in time on the order of hundreds of picoseconds (ps). This only begins to approach the lower end of spectroscopically observable data. Obviously a direct simulation of the folding of even a small protein is not feasible, unless additional assumptions and techniques (e.g. modifications of the model, potential function or method) are employed. For these reasons, it is not feasible to simulate a Ca^{2+} binding protein and hope to observe the dissociation of the Ca^{2+}. Rather the question is whether the molecular dynamics are being accurately simulated and whether the dynamical properties (e.g. conformational changes, flexibility) are being adequately sampled.

Empirical Potential Functions

The empirical potential function is an essential element of all computational chemistry methods described in this thesis. It's place in classical computational chemistry is analogous to that of the Schroedinger wave function in quantum chemistry - all the computed properties of the system depend on the potential function. The potential function is possibly the major theoretical limit to the accuracy one can obtain from a macromolecular simulation and its development and application result

from a compromise between the exact solution and approximations which make simulations feasible with current computational facilities. Thus over time, as computational facilities increase, the potential functions as well as other approximations used in macromolecular simulations will change.

The integral equations of statistical mechanics that relate the intermolecular potential, via pair correlation functions, to the macroscopic properties of molecular systems, are very complex; the exact theory can not be applied to complex systems (non-homogeneous) without large approximations. This complexity was initially a major motivation for applying computer simulation techniques to the study of molecular systems. These early simulations provided "experimental" evidence about the quality of approximate theories of some simple, monatomic, homogeneous systems.

At first, a system of argon atoms was simulated due to the relatively simple intermolecular potential of noble gas atoms. Later simulations were performed on homogeneous solutions of more complex molecules, such as water, whose potentials include large attractive terms such as electrostatic interactions.

The first potential applied in computer simulations of these simple systems was the hard sphere potential, defined as

$$v(r) = \begin{cases} \infty ; & r \leq d \\ 0 ; & r > d \end{cases} \qquad (7)$$

where v(r) is the potential between two atoms at a distance r and d is the atomic diameter. This model is applicable to solutions of argon atoms, where the potential is dominated by repulsion between the atoms' hard cores (resulting from overlap between their outer electron shells). Simulations using this potential function were remarkably accurate [Alder, 1957]. This model does not accurately represent a "true liquid state" as that is best described by a potential which includes an attractive term. This was first accomplished by a square-well potential, defined as

$$v(r) = \begin{cases} \infty ; & r < d \\ -e ; & d \leq r \geq d \\ 0 ; & r > d \end{cases} \qquad (8)$$

where d is a coefficient of the atomic diameter (typically 1-2) and e is the maximum attractive interaction.

At this point, the development of intramolecular potential functions diverged according to the complexity of the system under study. The definition of the accuracy of these potential functions also diverged. For simple, monatomic systems, such as inert gases, essentially exact results were obtained from a potential function for argon developed in 1971 [p168, Rigby, 1986]. Highly accurate potentials were also developed for small polyatomic systems such as H2, He, water and benzene. These functions, however, are truly optimized for accuracy and not computational feasibility. They have functional forms which are computationally intensive such as their inclusion of many-body terms. The field of modelling these small molecules will not be considered further. It is mentioned as an example of the accuracy that can be obtained in an empirical function.

The next step in the development of empirical potentials that apply to larger molecular systems was the Lennard-Jones potential, defined as

$$v(r) = 4e\left[\left(\frac{s}{r}\right)^{12} - \left(\frac{s}{r}\right)^{6}\right] \qquad (9)$$

where s is the collision diameter of two atoms and e is the depth of the potential well (maximum attractive interaction). This is the form in the empirical potential functions used today to represent the non-polar, non-bonded interactions between atoms. Comparing the hard-sphere to the Lennard-Jones potential function one can see an analogy to the relationship between the ideal gas law and the van der Waals equation of state (which was one of the first successful attempts to account for the major repulsive and attractive interactions in real gases). The exponents (12-6) used in the above equation were chosen as a compromise between ease of computation and accuracy.

Potential functions used in simulations of biomolecules do not include many-body terms. They often also do not include higher-order electrostatic interactions such as dipole-dipole interactions (polarization) which are forces that contribute significantly to the interaction potential of these systems, especially in a solvent environment. Instead, the parameters are modified to optimize the accuracy of the potential function within the forms used and the computational limits of the model. What results is a potential function which contains "effective" potential terms whose form resembles different types of physical forces (such as "electrostatic", "van der Waals", "bond stretching", etc). Thus no single term in the potential function models a single type of physical force. The mixing that results from the parameterization should be borne in mind when drawing conclusions from the results of molecular dynamics simulations - when these individual terms are analyzed for their "contribution" to the behavior of the system. Assigning effects to specific types of potential interactions from data on individual terms in the potential function can be misleading.

Pair-wise additive potentials developed on small molecules and molecular fragments chemically related to those found in proteins (e.g. carbonyl and carboxyl groups, phenyl rings, etc) are used in modelling macromolecular systems such as proteins. These potentials are parameterized to reproduce experimentally observable structures. When used in macromolecular simulations, however, often further computational approximations are made in order for the simulation to be feasible. In the computations performed in these studies, for instance, non-polar hydrogen atoms were not explicitly simulated (e.g. methyl groups were modelled as "extended" atoms) and the computation of the non-bonded (long range) terms in the potential function (i.e. van der Waals and electrostatic terms) was limited to those atoms within a "cutoff" (maximum) distance of each other.

These approximations, such as the non-bonded interaction cutoff, are often justified on theoretical grounds (e.g. electrostatic forces between non-polar atoms are

negligible at separations greater than 10 angstroms) but should always be tested to determine their effect on the accuracy of the simulation. The error introduced by the potential function and other approximations in the actual calculation of the trajectory can sometimes be measured and thus accounted for. Sometimes they can be partially overcome by simulating a larger ensemble of conformations (larger trajectory) if one is mainly interested in average properties instead of the microscopic trajectory.

Models for Treating Solvation Effects on Protein Structure and Function

The importance of solvent in accurately simulating proteins was emphasized even in the early molecular dynamics studies of proteins, and the inclusion of solvent effects remains a significant problem [Brooks, 1986; McCammon, 1987]. Solvent exerts enormous effects on the biophysical properties of proteins. It is essential to the native structure, dynamics and expression of activity in proteins [Edsell, 1983]. One example of the effect of solvent on the activity of proteins is the dependence of myoglobin activity on solvation; the rate of oxygen binding to myoglobin is dependent on the degree of solvation [Franks, 1979].

Understanding of the role of water, the solvent of biophysical systems, in maintaining the structural stability of proteins has been advanced by studies of xray crystallographic structures of proteins. A class of solvent molecules, called crystal waters, are observed in high resolution xray structures of proteins. These waters are tightly bound to the protein. They are observed forming hydrogen bonds that bridge pairs of charged side-chains. They also form networks (chains) of hydrogen bonded waters that can extend deep within the protein structure as well as occupying well defined solvation shells around residues on the surface of the protein. All these water molecules appear to be essential for the structural integrity and activity of proteins [Kuhn, 1992].

While the structural interactions of proteins with solvent molecules is observed from xray crystallographic studies, from which their role in the maintenance of structure is hypothesized, the relationship between solvent and protein dynamics is not well understood. The influence of solvent on protein dynamics is probably as significant as that on protein structure. Experimental data at the molecular level on dynamic interactions between proteins and their solvent surroundings is very difficult to obtain. General statements, such as the fact that protein internal mobility is reduced in the absence of solvent, have not lead to an understanding at a molecular level of the role of solvent in protein dynamics and activity. Molecular modelling studies can, potentially, increase our understanding significantly in this area.

As the area of solvation effects on protein dynamics has progressed the theoretical methods have shifted from the early continuum models of solvent to more discrete models. With the application of discrete solvation models more progress at the molecular level is possible than with continuum models. This is due to the molecular nature of discrete solvation models. These models include explicit solvent molecules in their representation of solvent interactions with proteins and allow observations of the specific molecular interactions that underly the observed macroscopic effects on the protein. At distances less than 10.0Å (the Debye distance for water) it is incorrect to treat water interacting with protein as bulk solvent but it can be appropriately modelled as a system of discrete molecules.

These explicit interactions between waters and the protein, along with the effect of the bulk solvent, produce the observed macroscopic effects of solvent on the protein. Solvent models which include these explicit molecular interactions may best describe and model the solvent. While other methods of including solvent in simulations of protein dynamics may do equally well in reproducing the structure and even dynamics of the solvated protein, the molecular basis for the solvent effect is probably beyond the

reach of these methods if the model excludes the explicit interactions of solvent with the protein.

In molecular dynamics simulations of globular proteins where solvent is not included there is an overall compaction of the molecule, observed as a reduction in the radius of gyration, where surface residue side-chains have been observed to form inappropriate hydrogen bonds and salt-bridges to nearby atoms on the surface of the protein [Levitt, 1982]. Overall structural changes may be small but the molecular dynamics of the protein can be significantly affected [Axelsen, 1988]. It is widely believed that inclusion of solvent effects is required to accurately simulate a solvent exposed region such as an active site or ion binding site.

Until recently protein simulations were generally performed in vacuo. Continuum models were developed partly as a result of the lack of computer resources to feasibly include the thousands of explicit waters needed to explicitly solvate even a small protein. Another problem of the discrete solvation methods was the lack of accuracy in the water molecule potential functions. The accuracy of potential functions used in protein simulations have increased as the field has progressed. These explicit solvent models have produced useful results and the periodic image boundary method of simulating solvated proteins is generally accepted as the most accurate method of simulating solvated molecular systems.

Boundary Methods for Simulating Solvated Proteins

To simulate a solute in an essentially infinitely large system of bulk solvent it becomes necessary to define boundary conditions that will truncate the representation of the system at some point without introducing arbitrary edge effects. The periodic image method provides the most nearly complete representation of solvent interactions and is regarded as the most accurate method in use. This method, shown in 2 dimensions in Figure 4, removes the system's edges by simulating an infinite periodic

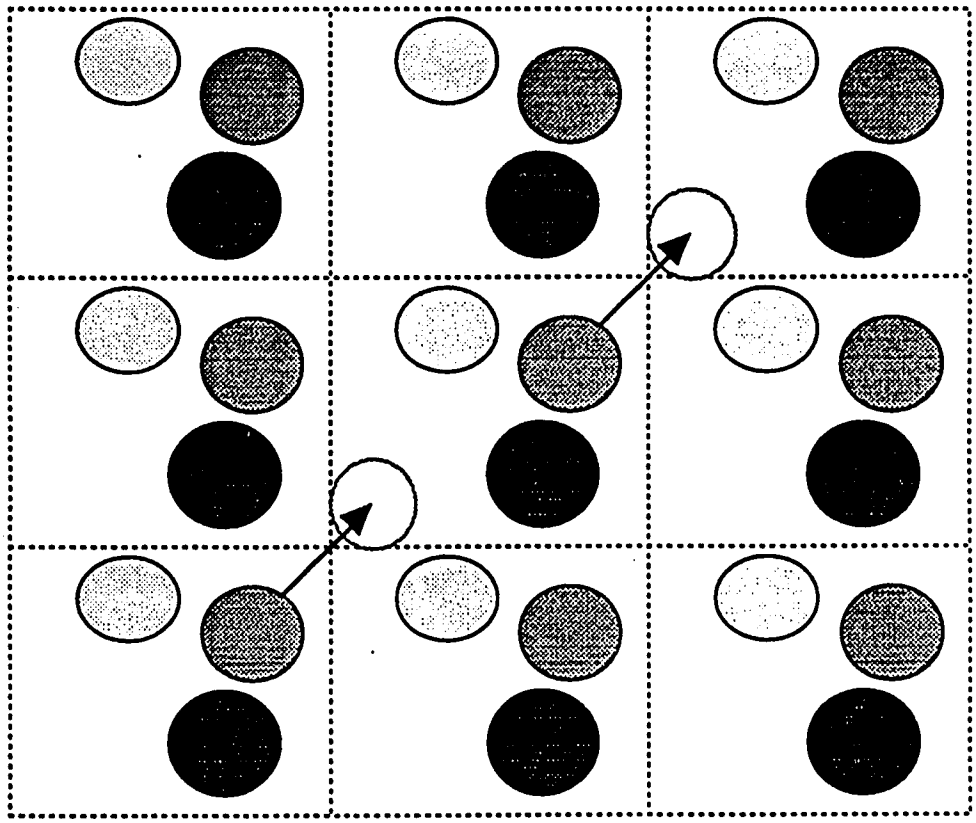


Figure 4. Two dimensional schematic of the periodic boundary method. A main cell is shown surrounded by images. As a molecule drifts out of the main cell a corresponding image molecule enters the main cell. The image molecule that is now in the main cell becomes the new position of this molecule used in the calculation (replacing the other copy of the molecule that becomes an image when it leaves the main cell).

array of identical repetitions (images) of the system. Interactions between atoms in the main cell and surrounding cells are included in the simulation. Any atom that drifts out of the main cell during the simulation, into a surrounding cell, is replaced by its image which, due to symmetry, is drifting into the main cell. This procedure practically eliminates edge effects but greatly increases computational demands by increasing the number of atoms explicitly included in the simulation and thus the number of atomic interactions that must be computed. For many systems (e.g. macromolecular enzyme-ligand complexes) it is not feasible to use periodic boundary conditions due to the limitations in computational resources.

The greatest advantage of simulating solvated proteins with the periodic image boundary method is its accuracy. It is the most widely used explicit solvation method in molecular dynamics simulations. Its primary disadvantage, that it is computational very expensive, limits its application to very small macromolecular systems. The other major limitation of this method is that it is only applicable to equilibrium systems. This excludes its use, for example, in simulations of chemical reactions. Also, for properties of the system with a frequency lower than the propagation of sound this method introduces an error arising from the propagation of information across the image boundaries which is then overlaid onto the equilibrium behavior of the system. Thus the model is not truly a purely solvated molecule but rather a highly solvated crystal.

For many interesting biochemical processes, such as enzymatic reactions, enzyme-ligand binding, and energy transport, the periodic image boundary method described above is not applicable. In these cases, the system would not be allowed to explore non-equilibrium behavior, such as the thermal relaxation involved in reaction processes. The structures of interest in the above examples often share the property of being spatially localized to specific "reaction zones" within proteins. Examples include active sites and ligand binding sites on enzymes. This property is exploited by boundary

methods which use reduced dynamics models of the macromolecular system to concentrate computational resources on the spatially localized region of interest. This approach often allows explicit solvent molecules, and their interactions with the protein, to be included in a computationally feasible simulation.

The scheme for partitioning a system into different regions, where each region is modelled differently, forms the basis of the reduced dynamics method. The level of complexity (e.g. the degree to which different regions of the system are included in the simulation) is variable and this flexibility is an added advantage of this method. This allows each model to be optimized according to the structure being simulated. Thus the degree to which explicit solvent is included may vary depending on whether a solvent exposed binding site or a deep active-site cleft is being studied. In each case the degree to which elements of the protein structure remote from the reaction zone are included can be also be varied, from very little for an isolated solvent exposed binding site to the complete protein for a central active site which induces large conformational changes upon ligand binding.

The reduced dynamics approach has been proposed for localized chemical processes, structures such as ligand binding sites on macromolecules, and macromolecular complexes for which it is inefficient (or not feasible) to use periodic boundary conditions [Brooks, 1983a, 1986]. Concentrating computational resources on modelling a localized region of a macromolecular structure is often the only feasible method of simulating these structures.

In the reduced dynamics methods the reaction zone is usually defined as a spherical volume. This part of the structure is modeled with the greatest accuracy (e.g. all degrees of freedom are included). Surrounding the reaction zone is a concentric shell called the buffer region. In the buffer region atoms may be restrained and/or additional conditions imposed (as described in Methods for the harmonic boundary method and the deformable stochastic boundary method). The molecular dynamics of the buffer are

perturbed by the boundary conditions imposed, however while these perturbations are unavoidable this region is by definition not the focus of the study. The rest of the structure (i.e. atoms outside the buffer region) is considered a "reservoir" and these atoms are completed constrained (i.e. no degrees of freedom) or excluded from the simulation.

Artifacts may be produced when the interactions between the reaction zone, buffer region and reservoir are not properly accounted for. When strong electrostatic forces are imposed on the buffer region a highly polarized, or "frozen", structure can be formed. This can result in strong hydrogen bonding interactions between the buffer solvent and atoms in the reaction zone producing a rigid buffer structure which perturbs the dynamics of the reaction zone [Brunger, 1984].

Classical molecular dynamics (MD) simulations were initially applied to protein systems *in vacuo*, without explicit solvent [Alder, 1959; Stillinger, 1974]. It is well known that the solvent environment of proteins has an enormous effect of the properties and activities of these molecules [Edsel, 1983]. The inclusion enough explicit solvent molecules around a protein in order to simulate a bulk solvent environment is not currently feasible. For this reason, until recently simulations of proteins were performed *in vacuo* or with only a few water molecules included. These water molecules were usually obtained from the xray crystallographic structure and are called crystal waters.

The model of the solvated proteins simulated in this thesis include a representation of the bulk solvent environment. Since bulk solvent is, by definition, a macroscopic phenomena, it has proven effective to apply potentials of mean force to model the interaction of bulk solvent with proteins. This potential is an integral part of the deformable stochastic boundary method described in the Methods section.

III. METHODS

While all the methods used in this thesis are from computational chemistry, quantum chemical effects are neglected; all the methods are based on classical mechanics. Quantum mechanical effects become significant when studying chemical reactions or small particles confined in microscopic spaces. Using classical mechanics in these studies is a valid approximation.

A. Molecular Dynamics Simulations

The simulations presented in this thesis were performed with the CHARMm package of software [Brooks, 1983b] using parameter set 19, the Ca^{2+} parameters developed in our group [Hori, 1988], and the TIP3 water model [Jorgensen, 1981]. The extended atom approximation was used (no explicit non-polar hydrogens were modelled). Polar hydrogens were modelled on all structures obtained from experimental methods, with the CHARMm program, using standard bond lengths and angles. The SHAKE algorithm [van Gunsteren, 1977] was used in the molecular dynamics simulations which allowed the timestep to be increased (doubled). The computations were performed on various computers, including those from SGI [Silicon Graphics, Inc., Mountain View, Cal.], Convex [Convex Computer, Inc., Dallas, Tex.] and Cray [Cray Computers, Inc., Minneapolis, Min].

The molecular graphics package QUANTA [Molecular Simulations, Inc, Waltham, Mass.] was used for graphical analysis of the molecular structures and the images were displayed on SGI graphics computers. The molecular graphics figures presented here were produced on a Tektronix RGB II [Tektronix, Inc., Wilsonville, Oregon] plotter from the QUANTA display on the workstations.

B. Molecular Dynamics Boundary Methods

Several boundary methods are included in protein simulations performed in this thesis. The boundary methods are an integral part of modelling a solvent environment. In the first study of the cyclic peptide-ion complex PG3 three boundary methods were used in order to compare the simulation results of two different reduced dynamics methods with the standard approach to solvation which is the periodic image boundary method. Besides this periodic boundary method the two reduced dynamics boundaries were the harmonic buffer method and the deformable stochastic boundary method. The specific values of the parameters and variables used in these methods are described with the study in the results section.

The periodic image boundary method is the most complete representation, of solvent, commonly used in molecular dynamics simulations. For this reason it has become the standard method used in the field. The periodic image boundary sets up virtual images of the system which surround the main system; all interatomic interactions between the images and the main system are included in the calculated potential energy. When an atom of the main system drifts out of the main cell (into an adjacent image cell) a symmetry image atom is drifting into the opposite side of the main cell; this image atom then replaces that atom, that drifted out of the main cell, becoming part of the main system. Thus the volume of the main system is constant and the surrounding images provide a boundary without any discontinuous edges. The procedure is computationally very expensive - prohibitively expensive for simulations of large proteins. Thus, while it accurately simulates a solvent environment another method was necessary to perform the simulations of the solvated subtilisin Ca^{2+} binding site B.

The harmonic buffer boundary simply constrains an outer buffer shell of solvent (in a spherical system) to a reference position (taken from the optimized

starting structure) using a harmonic spring potential. The form of the harmonic potential is

$$U_{HARM} = K(r-r_0)^2 \quad (10)$$

where K is the harmonic force constant and r_0 is the atom's reference position. The force constant is determined empirically; in order to minimize any perturbation of the reaction zone it should be the smallest value that will prevent significant mobility of the solvent in the buffer region (and provide a steric barrier). This basic form of the harmonic restraining force can be modified by, for instance, using a power higher than two (e.g. cube power of the displacement distance from the reference position).

Since the range of phenomena studied by these computational methods is continually growing, the utility of these boundary methods are considered in light of future applications. The deformable stochastic boundary method [Brooks, 1983a, 1985, 1986] models a physical system in which energy and mass can be transferred across a boundary, thus allowing the study of non-equilibrium processes such as chemical reactions and activated conformational changes. The deformable stochastic boundary (SBMD) represents a more complicated model of the solvent interactions and environment of the protein than does the harmonic buffer. It includes Langevin forces which model the Brownian motion of the solvent in the buffer region along with a potential of mean force which models bulk solvent effects on the explicitly included solvent molecules in the simulation.

In this method Langevin dynamics are introduced into the buffer region, to reproduce Brownian motion, via a stochastic mean force, F_r (random, thermal disorder) and a frictional force, F_d (viscous drag). These forces are defined as

$$F_r(t) = mf_r \quad (11)$$

$$F_d(t) = -m\beta v(t) \quad (12)$$

where $\langle \mathbf{f_r}(t) \rangle = 0$ ad $\langle \mathbf{f_r}(t) \cdot \mathbf{f_r}(0) \rangle = 6k_B T \beta \partial(t)$

(k_B is Boltzmann's constant, T is absolute temperature, β is the frictional coefficient and ∂ is the Dirac delta).

The SBMD method also includes a potential of mean force to represent the interaction of bulk solvent (not explicitly represented in the simulation) with the explicit solvent molecules included in the simulation. The force between an explicit solvent molecule in the simulation (at $\mathbf{r}_0$) and bulk solvent, $\mathbf{F}_B(\mathbf{r}_0)$, is obtained as

$$\mathbf{F}_B(\mathbf{r}_0) = \int \mathbf{F}_B(\mathbf{r}_0,\mathbf{r}_t)\, rho\; g(\mathbf{r}_0,\mathbf{r}_t)\, d\mathbf{r}_t \qquad (13)$$

where the integration range is over the volume of the reservoir within the non-bonded cutoff distance from $\mathbf{r}_0$ (*rho* is bulk density, $g(\mathbf{r}_0,\mathbf{r}_t)$ is the pair correlation function and $\mathbf{F}_B(\mathbf{r}_0,\mathbf{r}_t)$ is the force between two molecules at $\mathbf{r}_0$ and $\mathbf{r}_t$). In these reduced dynamics simulations whose regions are spherical, concentric volumes (i.e. the reaction zone, buffer region and reservoir) the mean force representing bulk solvent is radially directed toward the center of the reaction zone (i.e. the origin of the system).

PG3 and Erabutoxin Model Systems

A direct comparison of solvated molecular dynamics simulations of a small protein, erabutoxin b [Smith, 1988], was performed with and without the SBMD boundary conditions. This protein is small enough to be modelled with the boundary conditions without any boundary forces being applied to the protein atoms (only the surrounding solvent). In this study the indirect effects on the protein, resulting from boundary conditions on solvent, were investigated.

The stochastic boundary method, used in the subtilisin and erabutoxin studies, was first evaluated by direct comparison of this new boundary method to standard boundary methods used in molecular dynamics simulations. The SBMD boundary method was compared directly to the periodic boundary method, which provides the

most complete representation of a solvent environment, and a third, more simplistic, boundary method. This study was performed on a small, peptide-ion complex, cyclo$\left([\text{-L-Pro-Gly}]_3\right)_2$-$Ca^{2+}$ (PG3) [Kartha, 1982]. No significant differences between the periodic boundary method and the stochastic boundary method were observed in this study and thus these results provided the foundation for using the SBMD in the erabutoxin and subtilisin simulations.

C. Free Energy Perturbation Method

Calculating the statistical thermodynamic properties of macromolecular systems via computer simulations is a very powerful new approach to the study of protein-ligand interactions. The classical statistical mechanical approach to the equilibrium properties of such systems would be to calculate the partition functions of the various species involved in the binding (i.e. solvated protein ligand complex, isolated solvated protein, isolated solvated ligand). This computation involves calculating the Boltzmann weighted distribution of energies of all conformations of each species. Even in the classical mechanical limit using empirical potential functions this is an intractable problem.

A perturbation approach, however, is feasible and studies using computer simulations to calculate free energy differences of small perturbations in macromolecular complexes are under active development. Absolute free energies are not obtained from these perturbation simulations - only differences between two systems. In studies of protein structure/function relationships difference free energies are usually of greater interest than absolute energies. Of the two perturbation methods in current use (windowing and integration) only the windowing method was used and will be described. The same theory underlies both methods.

The central idea to perturbation theory is the partitioning of the potential energy function (described above) into two parts which correspond, in the studies on

subtilisin, to the native and mutant proteins. The perturbed potential function is defined as

$$V(L) = V_0 + LV_0 \tag{14}$$

where V_0 is the potential defined for the reference system (e.g. native protein) and V_1 is the potential of the perturbed (e.g. mutated) system. The above potential function corresponds, at L=0, to the native system, and, at L=1, to the fully perturbed system. V(L) need not be a linear function of L, the perturbation variable.

The Helmholtz free energy difference can be written as

$$\Delta A = -\left(\frac{1}{\beta}\right)\ln\left(Z_0 - Z_1\right) \tag{15}$$

where Z_0, the classical configurational partition functions of the reference system is

$$Z_0 = \int \exp\left(-\beta V_0\right) \tag{16}$$

integrated over all conformations of the system. With V(L) defined above,

$$Z_1 = \int \exp\left(-\beta\left(V_0 - V_1\right)\right) \tag{17}$$

and multiplying and dividing by Z_0 yields

$$\Delta A = -\ln\left\langle \exp\left(LV_1\right)\right\rangle_0 \tag{18}$$

which is the fundamental equation of thermodynamic perturbation theory. In the case of a true perturbation - where V_1 is very close to V_0 - it is possible to approximate the above equation as

$$\Delta A = -L\langle V_1\rangle_0 \tag{19}$$

Here the free energy change of a perturbation is given by the perturbation potential averaged over the distribution of conformations of the unperturbed system. Application of this equation to molecular dynamics simulations is called the windowing approach in free energy perturbation simulations.

To apply this equation to molecular dynamics simulations of proteins (e.g. to calculate the free energy change associated with a site specific mutation) two approximations must be made. First, since the ensemble of reference (native) conformations over which the perturbed potential is applied is necessarily limited to the trajectory calculated in the simulation adequate sampling must be done to yield statistical significant results. This finite ensemble of conformations is vastly smaller than the whole conformational space (referred to in the above equation) which would yield the exact result according to perturbation theory. Second, the perturbation of a single residue mutation in a protein is much larger than a perturbation as defined in the above theory. These perturbations in proteins can be performed as step wise increments from the native to the final mutation. Thus a series of smaller perturbations are done and the free energy change from each step is summed. The free energy difference actually calculated from perturbation simulations is defined as

$$\Delta A = \sum_{i=1}^{n} -L\langle V_1 \rangle_i \qquad (20)$$

These two approximations in the computational application of free energy perturbation theory, known as the sampling and convergence problems, are the major limitations in the accuracy of this method.

D. Representations of Protein Structure and Dynamics

The utility of any method of analysis is a function of its ability to clearly communicate the data it represents. Several important characteristics of protein representations are ease of calculation, clarity, ease of interpretation, and unambiguity. The choice of which method to use however depends on what one is exploring. Thus it is obvious that several methods may be necessary for analyzing a given set of data and what is represented by different methods may overlap to some degree. The methods themselves may be generally characterized as representing

structural or dynamic data. Representations of protein structure are generally much more advanced than representations of protein dynamics.

The key to the analysis and representation of macromolecular structures is understanding the underlying rational topology of the system. Proteins are unbranched polymers of amino acids. The primary sequence, as the underlying rational topology of a protein, provides the foundation for many representations of both protein structure and dynamics. This topology is used in virtually all the methods described below.

All the methods described below are two or three dimensional representations of protein structural (or dynamics) data with the sole exception of the molecular graphics representations from the QUANTA computer graphics package. It must be mentioned that developments in the field of computer graphics have greatly assisted the analysis of these computer simulations. Their utility has previously been observed in the field of xray crystallography where computer graphics have helped advance methods for the refinement of protein xray crystal data.

Two Dimensional Plots

Linear Distance Plots and Pseudotorsional Plots

The linear distance plot (LDP) [Liebman, 1985] provides a clear representation of protein secondary structure. An LDP value is calculated for each residue except the last three. The LDP value for residue i is the sum of the interresidue alpha carbon distances between residue i and the next four residues, i.e.

$$LDP(i) = \sum \mathbf{r}_{CA_n} - \mathbf{r}_{CA_i} \qquad (21)$$

where $\mathbf{r}_{CA_i}$ is the position of the alpha carbon of residue i and the summation is over n from i+1 to i+4.

The LDP value reflects local protein backbone conformation on the carboxy terminal side of residue i. Since residues in protein secondary structures such as alpha

helices and beta strands have regular repeating backbone conformations there is a characteristic value for each secondary structure. The LDP values have a range of ca. 18.0Å to 36.0Å; lower values indicate greater curvature of the backbone (i.e. helical structure) while extended backbone conformations (i.e. beta strands) have large LDP values.

The LDP is a good method for representing protein secondary structure and for comparing two different protein conformations. A difference of two LDP plots (DLDP) is constructed by calculating the difference of LDP values of homologous residues. In the case of two different conformations of a single protein this is done directly. When comparing homologous proteins the homologous residues may be mapped (e.g. by inspection or sequence homology).

While the LDP value is not unique (many different backbone conformations can produce the same LDP value), and it does not represent differences with great resolution, the plot is however a useful represention of protein secondary structure. The DLDP compounds the LDP's lack of precision and so a second method, the difference alpha carbon pseudotorsional plot (DCAT), is also used. This is a more easily interpreted representation of differences in protein backbone conformation. The CAT value is the pseudotorsional angle formed by the alpha carbons of four consecutive residues and is calculated for each residue (except the first residue and last two residues).

A simple method of representing the molecular dynamics of proteins is the root mean square (rms) atomic mobilities which are plotted (by residue usually) in a two dimensional graph called the RMS. These values are calculated from the coordinate trajectory. This representation is analogous to the graph of Debye-Waller temperature factors (B values) used by xray crystallographers to represent xray data on atomic mobilities [Willis, 1975]. Along with the RMS plot a graph of B values is calculated from the coordinate trajectory data assuming a simple relationship between coordinate fluctuations and B values;

$$B = (8/3)PI^2\langle u\rangle^2 \quad (22)$$

where <u> is the mean deviation from the average atomic position. In this equation coordinate fluctuations are assumed to be isotropic and harmonic. Difference plots of both the RMS and B values can be calculated from two homologous structures.

Three Dimensional Plots

Distance Matrices, Hydrogen Bond Matrices and Correlated Dynamics Matrices

A well known representation of protein tertiary structure is the alpha carbon distance matrix or contact plot (DM) [Liebman, 1985]. This is a three dimensional matrix representation of protein structure where the value of the element DM(ij) is the distance between the alpha carbons of residues i and j;.

where $\mathbf{r}_{CA_i}$ and $\mathbf{r}_{CA_j}$ are the positions of the alpha carbons of residues i and j, (23)

where $\mathbf{r}_{CA_i}$ and $\mathbf{r}_{CA_j}$ are the positions of the alpha carbons of residues i and j, respectively.

The matrix is square symmetric with the elements numbered by residue. Only half the matrix need be displayed and is displayed as a contour plot whose contour limits are chosen empirically. While elements of both secondary and tertiary structure are represented by this method the main strength of this method is its ability to clearly display tertiary structure. The secondary structure is observed near the main diagonal. Bulges in the lowest contour (<8.0Å), along the diagonals adjacent to the main one, represent close contacts and greater curvature of the backbone and are interpreted as helical structures and turns (or loops). Beta strands and other extended backbone conformations are displayed as thin lines where the lowest contour extends only one or two diagonals off the main. The difference DM (DDM) is constructed analogous to the

DLDP and DCAT as the difference between each homologous element of two DM matrices.

The display of protein secondary structure in the DM is not as clear as in the LDP. The DM does provide a good overall view of the domain and tertiary structures of a protein. Its precision is limited by the arbitrary choice of contours however this does not limit its utility. One choice of contours for a protein structure would be the limits 8.0Å, 15.0Å and 22.0Å. The first contour would display all DM values less than 8.0Å as one color (for instance) and has proved useful for discriminating secondary structures along the main diagonal. The second contour (15.0Å) is useful to display tertiary structure interactions; 15.0Å is approximately the largest contact distance between secondary structures (i.e. alpha carbons from two adjacent alpha helices). The largest contour limit is much less useful and particularly arbitrary although it is used to delineate domain structures.

Both the DM and the DDM only give a general outline or overview of the conformations of the proteins displayed. There is no representation of specific side-chain conformations or interactions. Interpretation of these plots is usually limited to the more obvious structural features such as close tertiary contacts (in the DM) and large tertiary structurally differences (in the DDM).

We developed a method of representing protein structure which overcomes some of the disadvantages of the DM while still representing most secondary and tertiary interactions [Factor, 1991]. Advantages of this method, the hydrogen bond matrix (HBM), are its clarity and ease of interpretation. It is a very clear and easily interpreted representation of almost all secondary structures as well as many tertiary structure interactions except hydrophobic interactions (e.g. helix packing). These plots are complimentary representations of protein structure and, since the HBM is a square symmetric matrix, it is displayed in the other half of a matrix display of the DM (or DDM).

The HBM is similar to the DM in having its elements numbered by residue. The HBM element ij displays the hydrogen bond interactions between residues i and j. The hydrogen bonds are classified, by type of donor and acceptor atoms interacting, as either main-chain/main-chain (MM), main-chain/side chain (MS) or side-chain/side-chain(SS). A backbone carbonyl oxygen hydrogen bond to a backbone amide nitrogen is thus type MM. The HBM displays the permutations of these hydrogen bond types (i.e. MM, MM+MS, MM+SS, MS+SS, MM+MS+SS or none). More than one hydrogen bond of a given type (between residues i and j) is only counted once; thus the HBM representation of MS as the permutation of hydrogen bonds between residues i and j means that there are one or more main-chain/side-chain hydrogen bond interactions between these residues and the representation of MM+MS means that there are at least one hydrogen bond of each type. Solvent hydrogen bonds to protein residues are displayed along the main diagonal as all (i,i) and (i,i+1) hydrogen bond interactions are excluded.

The HBM is a sparse matrix. When contrasted with the DM plot its clarity and precision make it a powerful complement to the representation of secondary and tertiary structure displayed in the DM. Since the HBM is so sparse two HBMs are also displayed in one matrix (one on each half of the plot) and compared directly.

The available methods of representing protein dynamical properties are not as advanced as those for representing protein structure. This reflects their intrinsic complexity as well as the fact that less is understood about the relationship of protein dynamics to structure and activity. Recently several groups of researchers have been presenting work on correlation dynamics within proteins and are developing a new method of representing protein dynamics data from molecular simulations [Ichiye, 1991; Harte, 1990]. In this thesis a representation of covariant dynamic motions is presented based on these new developments.

The representation used in this thesis is called the covariance dynamics matrix (CDM). It is as a square symmetric matrix where, again, the elements are the residues of

the protein structure (as in the DM and HBM). The value of an element, CDM(ij), in the matrix is the covariance between the position of alpha carbons of residues i and j from any ensemble (e.g. a molecular dynamics coordinate trajectory);

$$CDM(ij) = \frac{\langle D_i \cdot D_j \rangle}{\mathrm{sqrt}\left(\langle D_i \rangle^2 \langle D_j \rangle^2\right)} \qquad (24)$$

where $\langle D_i \rangle$ and $\langle D_j \rangle$ are the average position of the alpha carbon atoms of residues i and j,respectively. Alternatively another set of residue atoms may be used such as backbone or side chain atoms.

The CDM shows highly correlated motion among adjacent residues (along the main diagonal) as well as points of close tertiary contact (e.g. beta sheet interactions, helix-helix packing, disulfide bonds and ligand binding sites). In this respect it is similar to the DM plot. But the more interesting aspects of the protein dynamics are the highly correlated (or anti-correlated) motions which occur between residues in the absence of any close structural interaction and it is here that this method can be most useful; in displaying dynamics properties that are not directly (i.e. clearly) represented by any other method commonly used.

IV. RESULTS

A. Comparison of Boundary Methods in Simulations of PG3

To compare the performance of these methods in the study of a Ca^{2+}-binding system, we simulated the molecular dynamics of the cyclic hexapeptide cyclo$\left([\text{-L-Pro-Gly}]_3\right)_2$-$Ca^{2+}$, termed PG3. The structure used in this study was the 2:1 complex with Ca^{2+} that this peptide forms in solution [Kartha, 1982]. The results of molecular dynamics simulations using three different boundary conditions were compared with respect to effects on the structure and dynamics of the peptide-ion complex, the Ca^{2+} coordination structure, and the structure of the solvent.

The three boundary methods that were compared are the standard periodic image (BOX), the deformable stochastic (SBMD) and the harmonic buffer (HAR) boundary methods (described in the Methods section). The results of the SBMD method were not significantly different from those of the standard BOX method. The HAR method results in changes in the structure of the peptide, the coordination of the calcium ion, and the solvent structure, when compared to the other two boundary methods. The dynamics results of the peptide-ion complex do not appear to differ among the three boundary conditions imposed which may reflect the rigid character of the cyclic peptide structure and the strong binding interactions of the peptide ligands to the Ca^{2+}. This study of the behavior of a model system under these three boundary conditions was the basis of the use of SBMD as an alternative to the standard periodic image boundary method. The SBMD method is computationally much more feasible for use in simulations of large macromolecular systems.

Model System

The peptide-ion complex (PG3) chosen for this study is small compared to other macromolecular Ca^{2+}-binding systems and has been the subject of previous studies by our group [Hori, 1988; Sussman, 1989]. The structural characteristics of the peptide have been determined by NMR spectroscopy and physicochemical studies have characterized its ion binding properties [Kartha, 1982]. Since this complex has been well characterized both experimentally and with the computational methods used in our laboratory it thus provides a good model system for this comparison of the effects of these different boundary methods.

The ion binding properties of this peptide are unusual as it binds Ca^{2+} and Mg^{2+} in two distinct structures; Mg^{2+} in a 1:1 complex and Ca^{2+} in a 2:1 complex with the ion sandwiched between two stacked peptides. The Ca^{2+} bound 2:1 complex was the structure studied here. The larger desolvation energy involved in binding Mg^{2+} is reflected in the solvent exposure of the bound ions, with the Mg^{2+} being more solvent exposed in its 1:1 complex than the Ca^{2+} in its 2:1 complex, and is the basis of an hypothesis for the ion selectivity expressed by the various solvent complexes of the peptide - the "Ca^{2+} selectivity" of the 2:1 complex [Sussman, 1989]. That hypothesis serves as a model for the mechanism of Ca^{2+} selectivity of other intracellular enzymes and Ca^{2+} binding proteins.

MD simulations were performed on the solvated peptide-ion complex cyclo$\left([\text{-L-Pro-Gly}]_3\right)_2$-$Ca^{2+}$ (PG3). The starting conformation of the peptide-ion complex, which is very similar to the structure determined by NMR, was obtained from a previous simulation which included 200 water molecules [Sussman, 1989]. The 200 explicit water molecules in that structure were included in the simulations performed in this study. The total number of explicit solvent molecules included in these simulations depended on the boundary method applied (see below).

Boundary Methods and Molecular Dynamics Simulations

The structure obtained from a previous molecular dynamics simulation [Sussman, 1989] was centered on the Ca^{2+} and used as the starting structure for applying all three boundary conditions.

The first boundary condition used was the standard periodic image boundary method. The simulation of PG3 using this method was called BOX. A 30.0 Å box of 1000 thermally equilibrated water molecules was overlaid and any overlaid water in close contact with other heavy (non-hydrogen) atoms (at a distance < 2.8 Å from an oxygen atom of the overlaid water) was removed. This resulted in 684 waters being overlaid on the starting structure for a total of 884 explicit waters. Images were constructed and the system optimized briefly with 100 steps of conjugate gradient minimization. This structure was used to start the molecular dynamics simulation with the periodic boundary conditions.

The BOX simulation used the default 7.5 Å cutoff distance for non-bonded potential energy interactions and a 8.0 Å cutoff for non-bonded list generation. Both the non-bonded list update frequency, and the image update frequencies were 10 steps. Bond lengths to polar hydrogens were fixed with the SHAKE algorithm and an extended atom representation was used for non-polar hydrogens. The Verlet algorithm for molecular dynamics implemented in CHARMm was used. The simulation was performed with a time step of 0.5fs after a larger time step of 1.0 fs was found not to conserve total energy. The system was slowly heated from 0°K to 300°K (2.0°K increments every 40 steps) and then equilibrated at 300°K with +/-10° windows for temperature scaling. Coordinate and velocity sets were recorded every 20 steps during the simulations and average structures were calculated for 20ps intervals.

The second boundary condition included in this study was a reduced dynamics method [Brooks, 1986] in which a harmonic restoring potential is used to restrain

water molecules in the buffer region and thus provide a steric barrier to the diffusion of solvent from the reaction zone. The starting conformation for this simulation, called HAR, was obtained by a procedure identical to that described for the BOX simulation above except that a 15.0 Å sphere of waters was overlaid (and the boundary conditions applied were different). No images were constructed and the resulting structure contained a total of 350 water molecules. A harmonic restraining potential was applied to oxygen atoms of waters in a 3.0 Å solvent buffer region (defined as the shell of waters 12.0-15.0 Å from the Ca^{2+} at the origin). The harmonic potential force constant applied was 0.01 kcal/mole and the reference position was the atom's position in the starting structure. Heating and equilibration were performed as described above for the BOX simulation. The HAR simulation was performed with a timestep of 0.5fs; as in the BOX simulation, the total energy was not conserved with a larger timestep of 1.0fs. Other parameters and variables used in this simulation were the same as in the BOX simulation and, again, coordinate and velocity sets were recorded every 20 steps and average structures were calculated for 20 ps intervals.

The third boundary method (second reduced dynamics) included in this study was the SBMD method of Brooks and the simulation was called SBMD [Brooks, 1983a]. The radial distribution function for the water molecules (TIP3 model) and the program to generate the stochastic boundary potential was kindly provided by Dr. C. L. Brooks, III. The starting conformation for the SBMD simulation was obtained by a procedure identical to that described above for the HAR simulation (i.e. PG3 solvated with 350 explicit water molecules). However, the boundary forces imposed on this starting structure were different.

Instead of the harmonic restraining potential, a potential of mean force was imposed on the water molecules explicitly included in the simulation. The potential of mean force, representing bulk solvent effects on the explicit waters in the simulation, modelled only oxygen-oxygen van der Waals interactions; only the van der Waals

potential was included due to problems in modelling other forces (e.g. electrostatic) between the bulk solvent and the system [Brunger, 1984]. A 3.0 Å buffer region (12.0-15.0 Å from the Ca^{2+}) was constructed; a frictional coefficient of 62.5/ps was applied to solvent oxygens in this buffer region and the region was modelled as a heat bath at 300°K with Langevin dynamics [McQuarrie, 1988]. No heating phase was performed; the molecular dynamics simulation began immediately with the equilibration phase at a temperature of 300°K. Other parameters and variables used in this simulation were the same as in the BOX simulation and, again, coordinate and velocity sets were recorded every 20 steps and average structures were calculated for 20 ps intervals.

Conservation of Total Energy and Temperature Equilibration

Total energy was conserved in the HAR and BOX simulations; fluctuations in the total energy (due to numerical roundoff error and the numerical methods of solving analytical equations of motion) was less than 0.01% for both these simulations. In the SBMD simulation the total energy is not constant as the temperature is constrained by periodic scaling of the atomic velocities. Figure 5 shows the behavior of the temperature during the simulations with the three different boundary conditions. The full dynamics simulation (BOX) has the smallest temperature fluctuations. From statistical mechanical theory [McQuarrie, 1988] temperature fluctuations in an equilibrium system are inversely proportional to the square root of the number of particles in the system. Thus the BOX simulation (with 884 explicit waters) would be expected to have much smaller fluctuations in its temperature than the two reduced dynamics simulations (with 350 explicit waters).

An analysis of the fluctuations in the individual potential energy terms of the two reduced dynamics simulations (HAR and SBMD) was done. In the SBMD simulation the large temperature fluctuations obviously resulted directly from the periodic scaling of atomic velocities done to simulate a heat bath. In the HAR simulation the constraint

Fig. 5a: BOX Simulation Temperature Plot

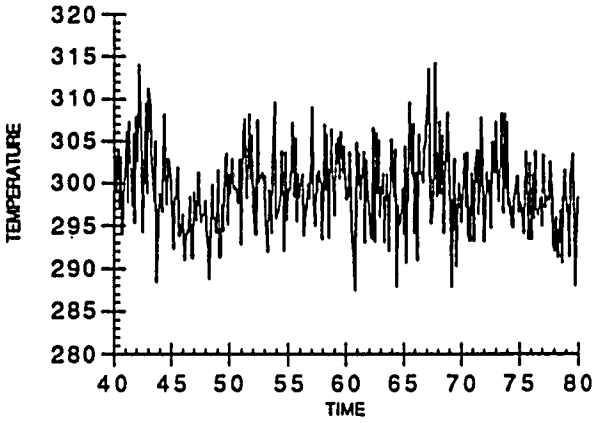


Fig. 5b: SBMD Simulation Temperature Plot

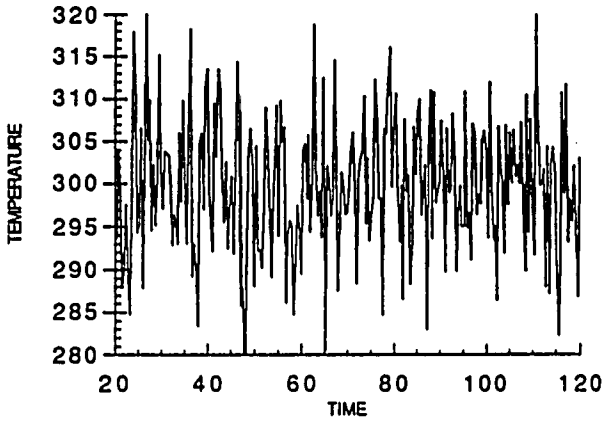


Fig. 5c: HARM Simulation Temperature Plot

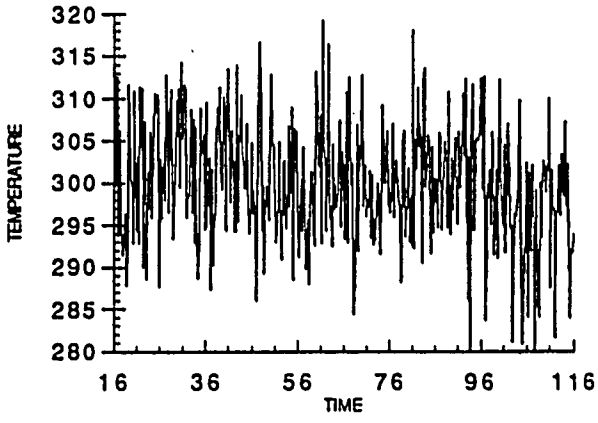
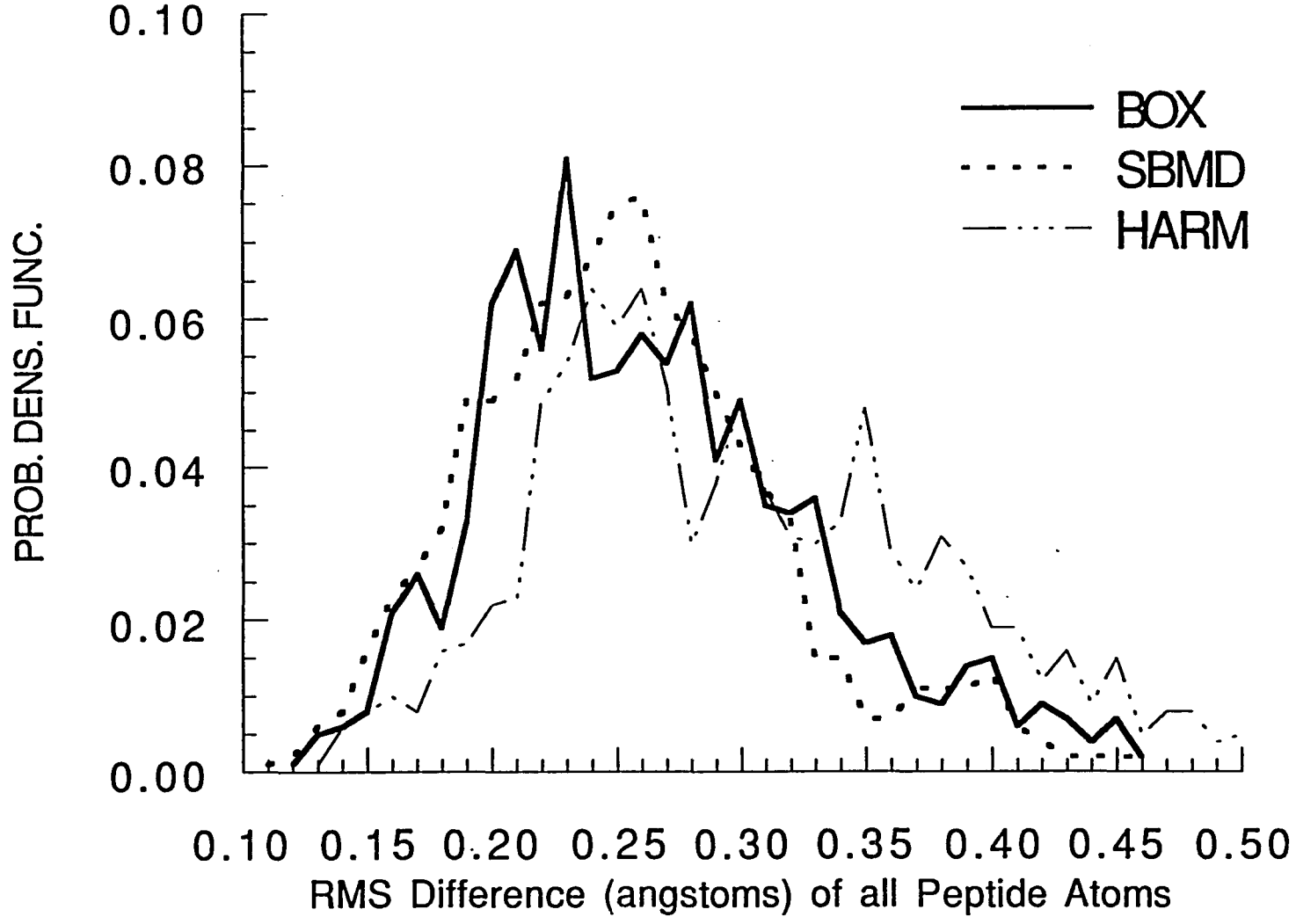


Fig. 6: Distributions of RMS Differences Between Last 20psec Average Peptide Structure and the Coordinate Trajectory Interval



potential on the buffer solvent molecules is 70% of the magnitude of the kinetic energy fluctuations. This suggests that the constraint potential on the buffer solvent contributed significantly to the large temperature fluctuations in that simulation.

Peptide Structure

Table 2 shows the rms difference between each 20ps interval average structure, obtained from the simulations, and the starting structure.

The average BOX peptide structure from the first 20ps interval differs from the starting structure by ca. 1Å, and this value does not change significantly during the simulation. This result suggests that the only significant change in the peptide conformation occurred during the optimization of the starting structure prior to the molecular dynamics simulation.

The average HAR peptide structure from the first 20ps interval also differs from the starting structure by ca. 1Å. However during the HAR simulation this difference increases to an rms difference of 1.28Å which is the largest difference from all three simulations. Analysis of the coordinate trajectory from the HAR simulation reveals a change in the peptide structure occurring between the second and third 20ps intervals. During this interval, the psi backbone torsional angles of Pro3 and Pro5 in the second cyclic peptide in the complex, labelled PG3B, change from ca. -15° to +130°. From the LDP and DM plots discussed in detail below, the backbone of PG3B is found to adopt a more extended, planar ring conformation.

During the SBMD simulation the rms differences from the starting structure decrease after the first 20ps of the simulation. The first average structure has an rms difference of 1.23Å which reduces to 1.15Å by the end of the simulation.

The peptide structure changes very little during the BOX simulation while during the HAR simulation the structure diverges from the starting structure. In the SBMD simulation the structure converges toward the starting structure. These data indicate

Table 2. RMS(CA) matrices of positional differences from the starting structure for average structures from 20ps intervals from the PG3 simulations.

Matrices of RMS Differences(Å) Between Average Structures[a]

BOX Average Structures

Trajectory Intervals	A	B	C	D	E	F
A	0.00					
B	0.19	0.00				
C	0.26	0.21	0.00			
D	0.34	0.29	0.11	0.00		
E	0.36	0.33	0.16	0.08	0.00	
F	0.36	0.31	0.16	0.10	0.07	0.00
Starting Crds	0.97	1.03	1.01	1.02	1.03	1.04

SBMD Average Structures

Trajectory Intervals	A	B	C	D	E
A	0.00				
B	0.26	0.00			
C	0.30	0.42	0.00		
D	0.28	0.42	0.17	0.00	
E	0.35	0.33	0.43	0.31	0.00
Starting Crds	1.23	1.30	1.17	1.11	1.15

HARM Average Structures

Trajectory Intervals	A	B	C	D	E
A	0.00				
B	0.28	0.00			
C	0.77	0.71	0.00		
D	0.85	0.86	0.35	0.00	
E	0.85	0.82	0.26	0.27	0.00
Starting Crds	1.00	1.03	1.24	1.30	1.28

[a] For BOX and SBMD, A to E indicate consecutive 20-psec trajectory intervals.
[a] For HARM, A to F indicate consecutive 10-psec trajectory intervals.

the robustness of the SBMD simulation as it is the only one that partially overcomes a conformational change in the peptide that occurred during the pre-dynamics optimization procedure and converges towards the starting structure. These results also suggest that the abrupt start of the SBMD molecular dynamics simulation with the

equilibration phase (no heating phase in used in this method) causes a perturbation in the system which is then partially overcome during the equilibration and production phases of the simulation.

Table 3 shows the rms difference between the final average peptide structures obtained from each of the simulations and the starting structure. The SBMD simulation results in a peptide structure which is about as close to the BOX simulation (rms difference 1.09Å) as to the starting conformation (1.15Å) while the HAR simulation differs more from these structures (1.28Å and 1.37Å, respectively). The results of the HAR simulation are thus less accurate than those of the SBMD simulation when compared to both the starting structure and the results of the standard periodic boundary method (BOX simulation).

Table 3. RMS(CA) matrix of positional differences among the last 20ps average structures from the PG3 simulations and the PG3 starting structure.

Matrices of RMS Differences(Å) Between Final Average Structures

Boundary Method	BOX	SBMD	HARM	Starting Crds
BOX	0.0			
SBMD	1.09	0.0		
HARM	1.37	1.26	0.0	
Starting Crds	1.04	1.15	1.28	0.0

Figure 7 shows the linear distance plots (LDPs) of the last average peptide structures and the starting structure, while Figure 8 shows the alpha carbon distance matrices (DM) of these structures. The LDP plots of these cyclic peptides were calculated in a modified way due to their cyclic structure; an LDP value could be calculated for each residue of the peptides as the last residue is bonded to the first. Separate LDP plots are shown for each of the two cyclic peptides in the peptide-ion sandwich and are labelled

Fig. 7a: Linear Distance Plots of Peptide Segment P3GA from Last Average Structures

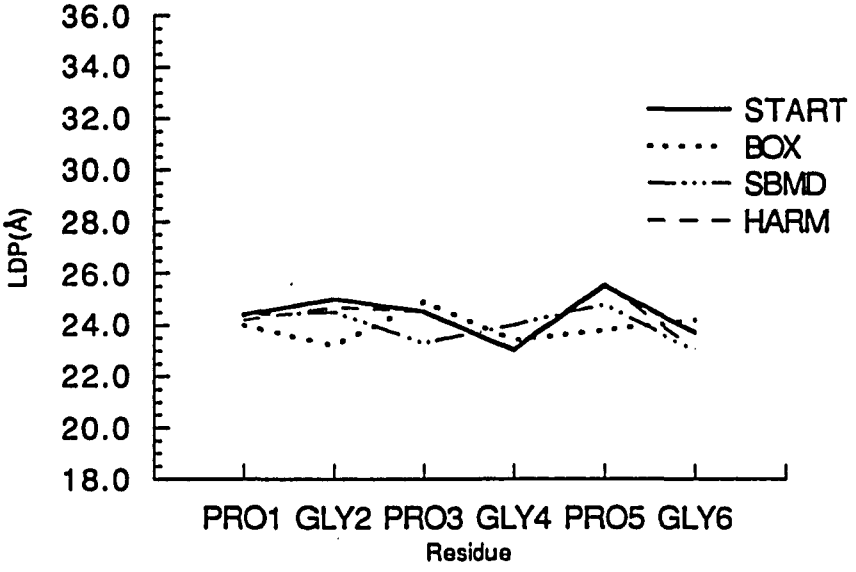
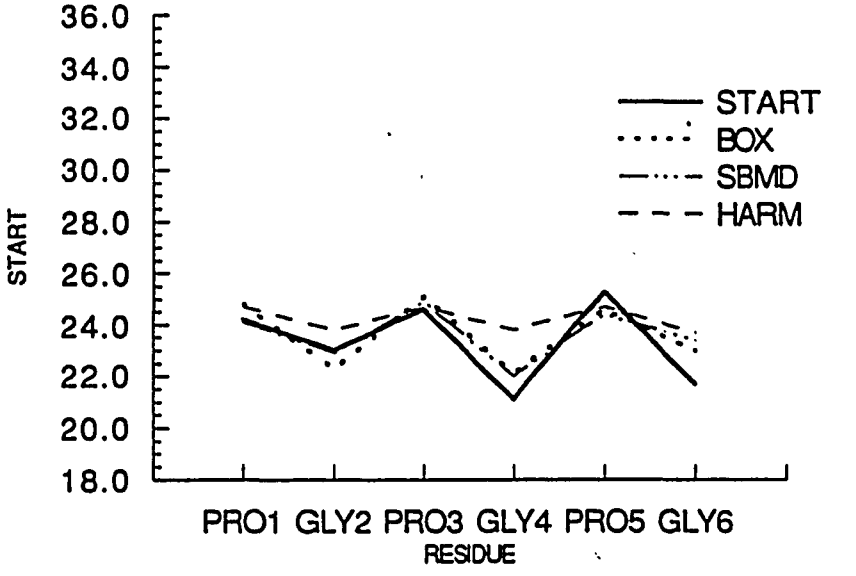


Fig. 7b: Linear Distance Plots of Peptide Segment PG3B from Last Average Structures



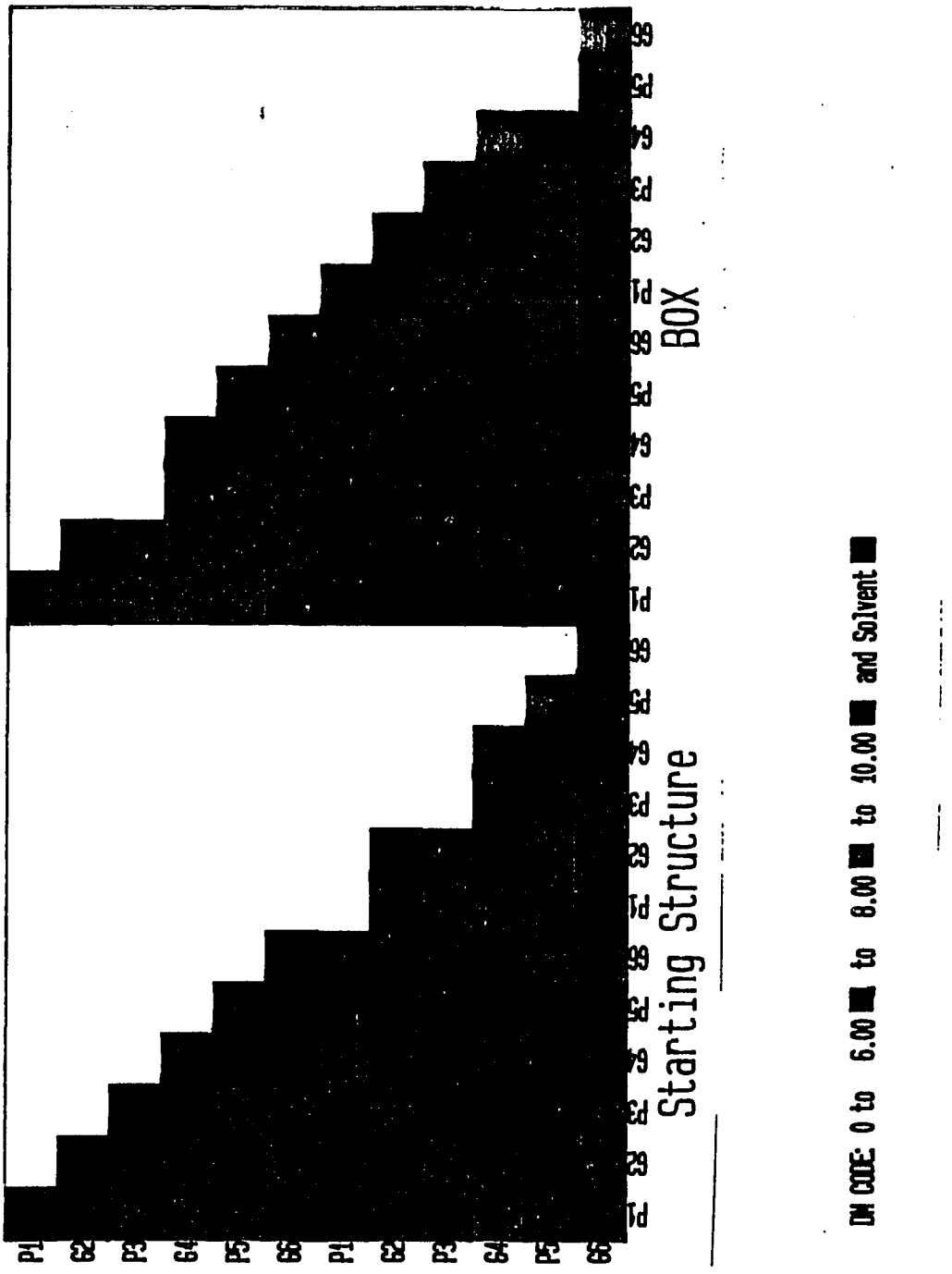


Figure 8a-b. DMs of PG3 starting structure (left plot) and last 20ps average structure from BOX simulation (right plot). Axes are peptide residues (P is Proline and G is glycine) with first six residues from first hexapeptide and second six residues from second hexapeptide in PG3 structure. Contours of DM (in angstroms) are 0-6(green), 6-8(orange), and 8-10(red). Residues with solvent within 3.3Å are indicated by a black square in the main diagonal.

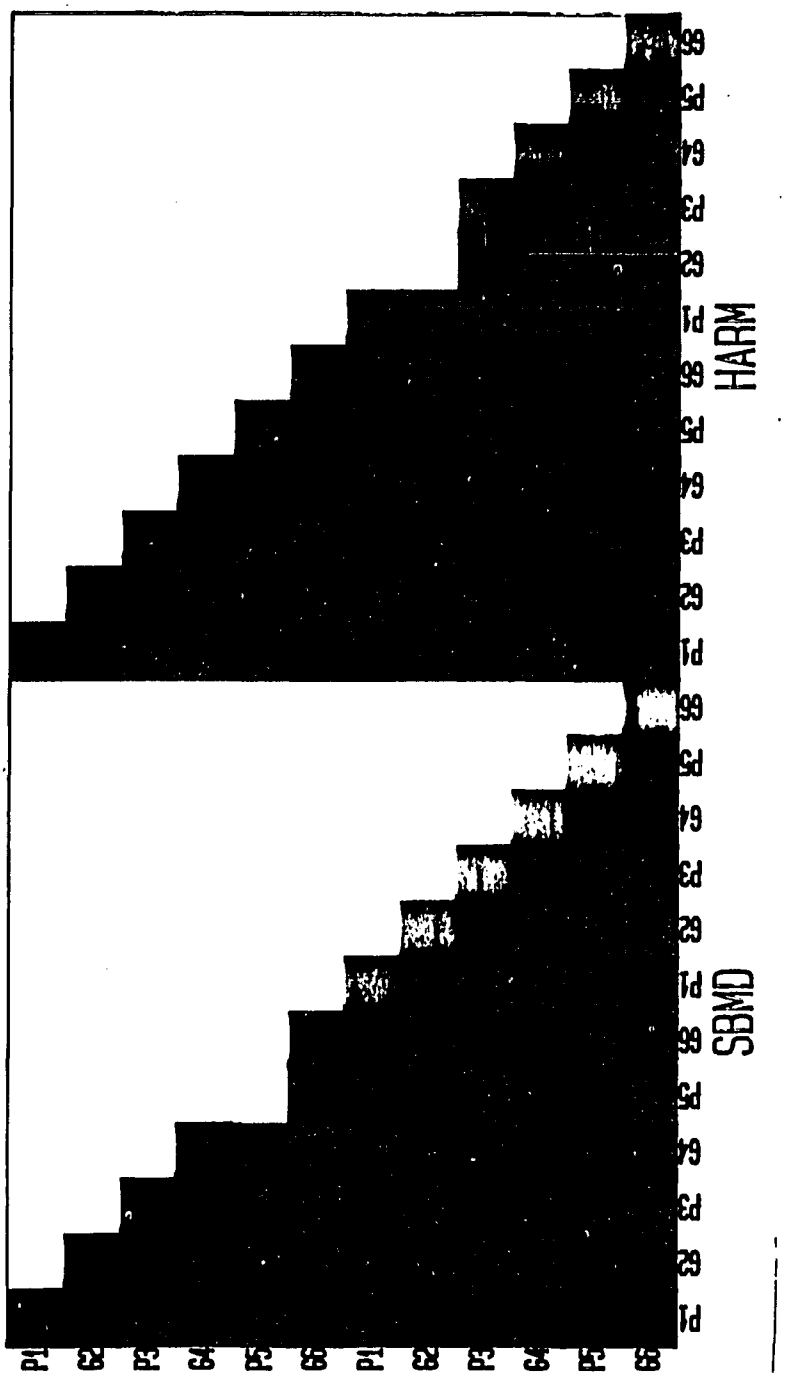


Figure 8c-d. DMs of last 20ps average structures from SBMD (left plot) and HARM (right plot) PG3 simulations. Axes are peptide residues (P is Proline and G is glycine) with first six residues from first hexapeptide and second six residues from second hexapeptide in PG3 structure. Contours of DM (in angstroms) are 0-6(green), 6-8(orange), and 8-10(red). Residues with solvent within 3.3Å are indicated by a black square in the main diagonal.

PG3A and PG3B as in the previous study [Sussman, 1989]. In these plots, P refers to proline and G to glycine residues (e.g. P1 is proline residue 1). Similarly, the DM plots are labelled with double sets of residues P1 to G6 where the first set is from peptide PG3A and the second from PG3B (reading across the x-axis and down the y-axis).

The conformational change that occurs in the HAR simulation around the G4 residue of the PG3B peptide can be observed in the LDP plot; as the HAR peptide undergoes its conformational change toward a more extended structure, the LDP values increase when compared to the LDP values of the PG3B peptide in the BOX and SBMD structures.

The LDP plots show very little difference between all three peptide structures and no regular secondary structure is seen. This reflects the rigid structural properties of the cyclic hexapeptide. While the LDP plots display secondary protein structure only, the DM plots display both secondary and tertiary structure. Thus the conformational differences between these structures are more easily observed in the DM plots. In Figure 8, the distances between residues within the PG3A peptide are shown in the upper left corner, while those of the PG3B peptide are shown in the lower right corner; the first P1 to G6 along the axes represent the residues of the PG3A peptide and the second P1 to G6 represent PG3B. The contacts between the two peptides are shown in the lower left corner of the plots.

The conformational change in the PG3B cyclic peptide of the HAR structure can be seen along the lower half of the main diagonal, in the DM plot, which displays a single thin green line of contacts representing the extended backbone conformation of this structure. The DMs of the starting structure and the BOX and SBMD structures all have multiple off-diagonal close contacts (< 6.0Å) which represent a different conformation of the PG3B peptide backbone (larger curvature and more compact structure) compared to the HAR peptide structure. The DM plot of the HAR peptide structure also displays a single unbroken line of close contacts (< 6.0Å) that extends

antiparallel from the main diagonal to the bottom left corner; these are interchain contacts showing that the two peptides, PG3A and PG3B, have moved closer together than the other structures and suggests a change in the Ca^{2+} coordination of this structure (discussed below).

A difference in the PG3A peptide structure can be observed in the BOX and SBMD structures as the two simulations result in a more planar conformation compared to the starting structure. However, once again the BOX and SBMD simulations results are similar and different from the results of the HAR simulation.

Peptide-Ion Coordination

The distances of the peptide carbonyl oxygens to the Ca^{2+} in the final structures from these simulations are shown in Figure 9. The starting structure along with the BOX and SBMD structures have eight oxygens coordinating the Ca^{2+} whereas the HAR structure has a coordination number of nine. The usual coordination number of calcium ranges from 6 to 8 [Martin]. The single solvent coordinating the Ca^{2+} in the starting structure at 2.8Å has become more closely boumd in the HAR structure (2.5Å). This water has diffused away in the coordination structures observed from the BOX and SBMD simulations (in which no waters coordinate the Ca^{2+}). Along with this change in the Ca^{2+} coordination, the structure observed in the HAR simulation is also the only one to result in a different coordination number (9) from the starting structure. Coordination number is a significant characteristic of Ca^{2+} bound complexes and this change represents a significant difference between the HAR simulation and the results of the other two simulations.

Figure 10 displays the probability density functions of the distances of the peptide oxygens coordinating the Ca^{2+}, during the simulations, and clearly shows another difference between the results of the HAR simulation and the SBMD and BOX

Fig. 9: Peptide Ligand Distances to Ca+2 in Final MD Structures

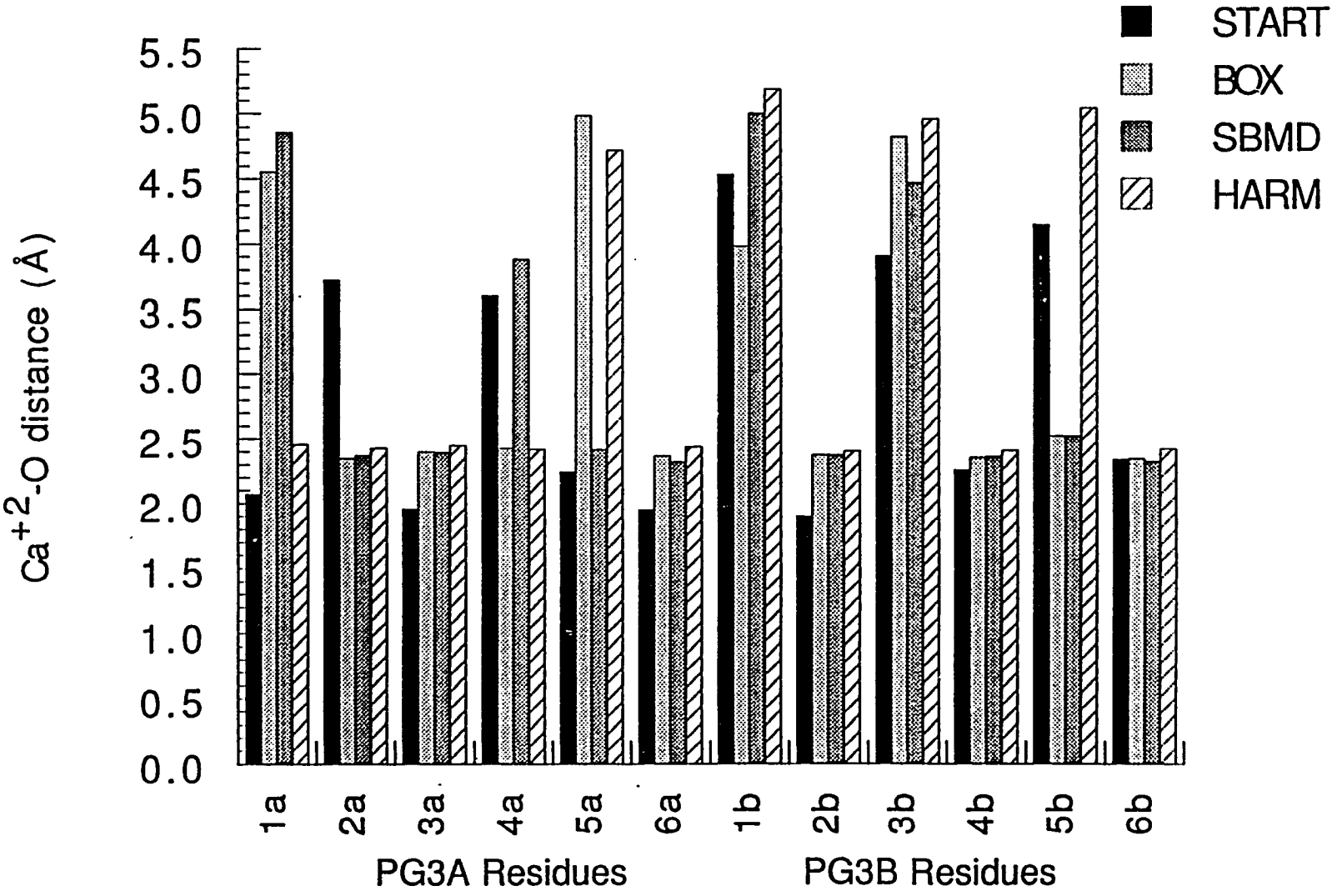
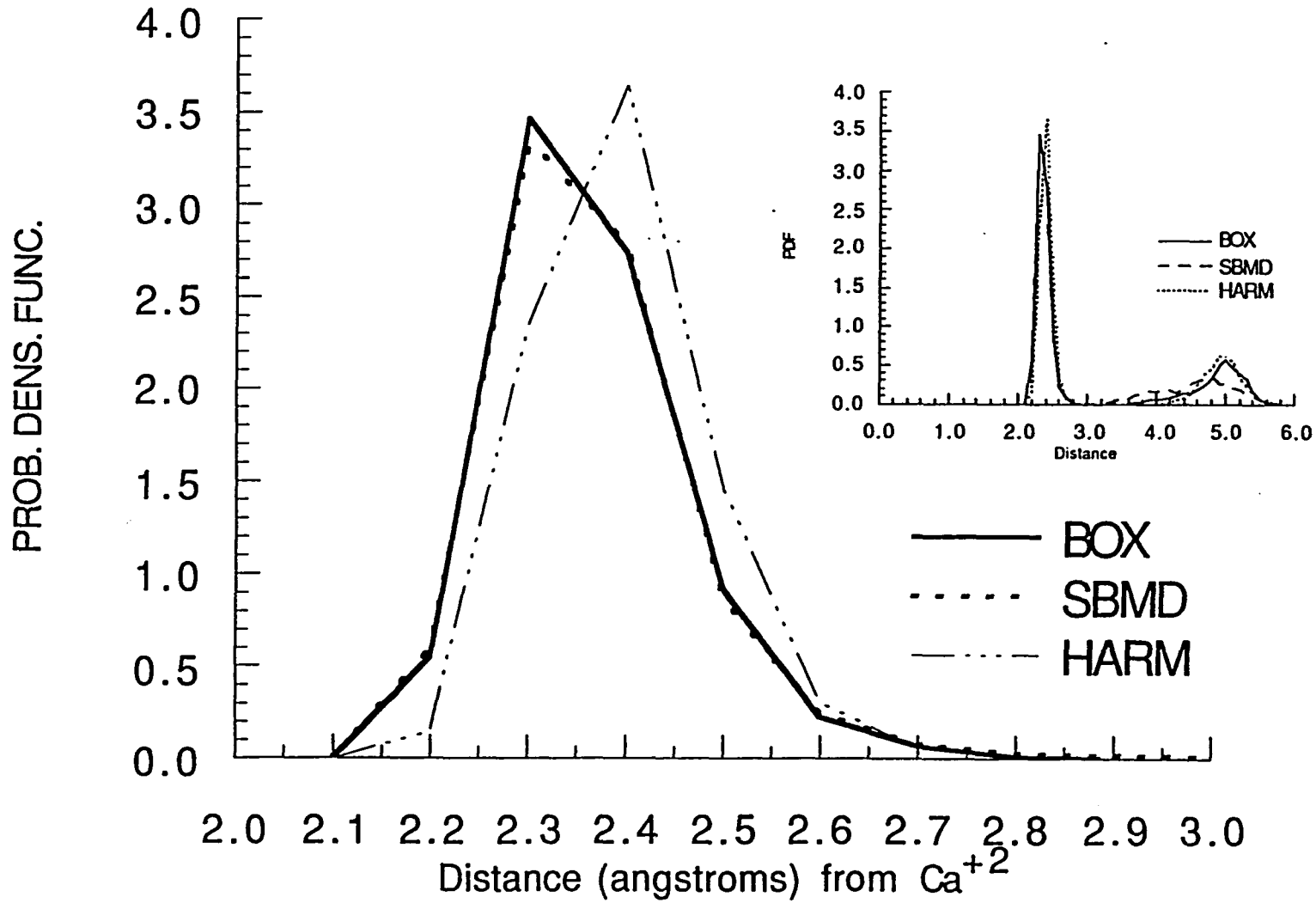


Fig. 10: Average Peptide Ligand (Carbonyl Oxygens) Structure Around Ca+2



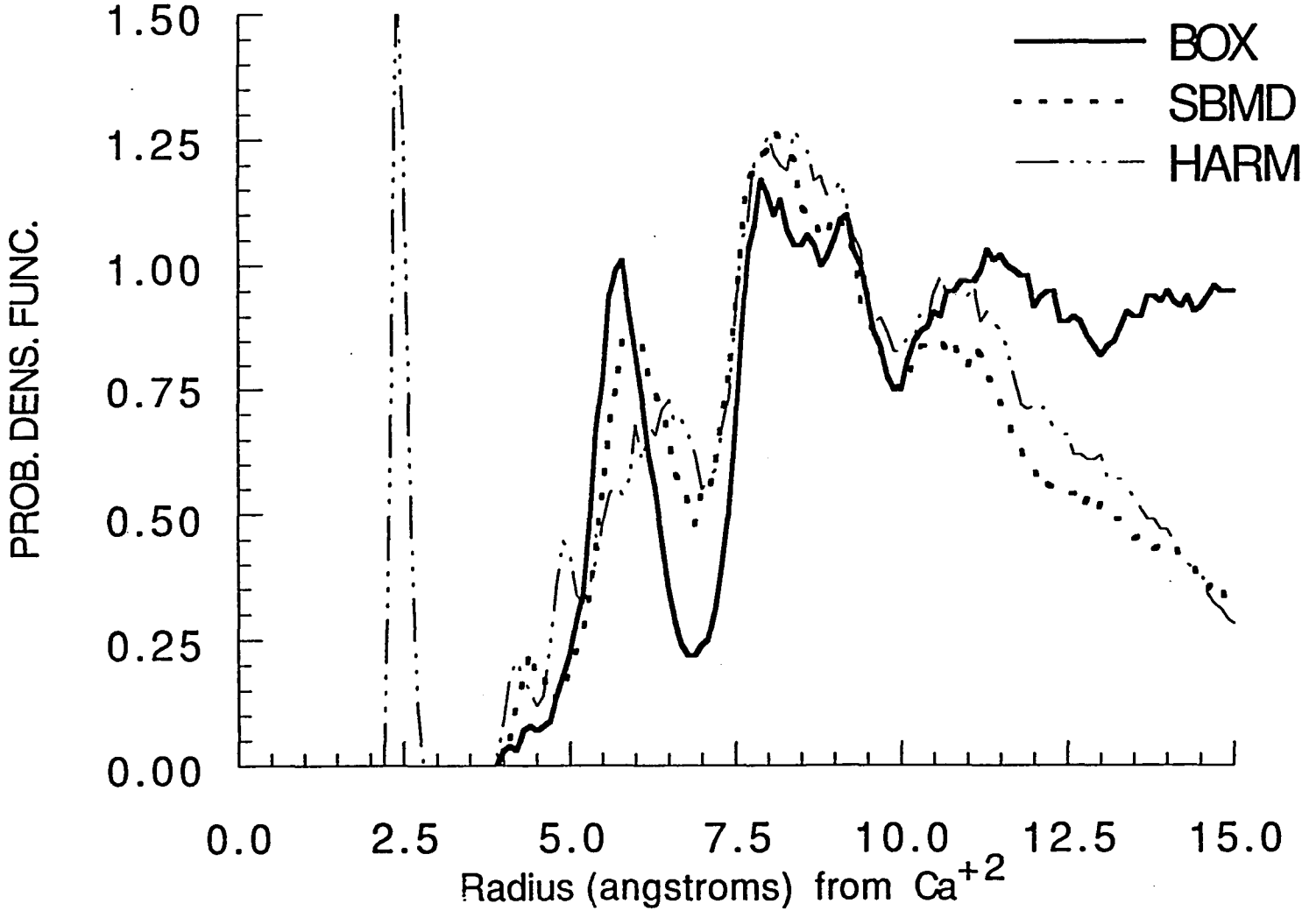
simulations; the average distance of the coordinating ligands to the Ca^{2+} is larger in the HAR simulation.

The SBMD simulation also results in a coordination structure of the PG3A peptide which is the closest to the starting structure. There is only a single switch of coordination between carbonyl oxygens of two adjacent residues (P1 to G2). The BOX structure is very similar to this; in it this switch and another from P5 to G4 is observed. Both the BOX and SBMD structures add a coordinating ligand from the PG3B peptide, the carbonyl oxygen of proline 5, to the Ca^{2+}. This ligand probably replaces the solvent ligand which was present in the starting structure. The HAR structure is significantly different from these other two structures with a total of five out of six carbonyl oxygens in PG3B coordinating the Ca^{2+} (in the structure observed from the HAR simulation).

Solvent Structure

The radial distribution functions of solvent around the Ca^{2+} (average solvent structure) of the three simulations is shown in Figure 11. Only the reaction zone data in the two reduced dynamics simulations, SBMD and HAR, are considered (up to 12.0Å from the Ca^{2+}). The lone solvent ligand of the HAR simulation is seen as a sharp peak at 2.5Å. At distances less than 4.0Å from the Ca^{2+} the peptide structures produce an excluded volume so no solvent is observed. The BOX and SBMD simulations have nearly identical first solvation shells with peaks at 5.5-6.0Å. The first peak of the HAR simulation is a much smaller, broader peak at 6.5Å representing a more disordered solvent structure and suggesting that in this simulation fluctuations in the peptide structure are larger than in the other simulations (see Molecular Dynamics below). All three simulations have a second peak (second solvent shell) at 10.0Å and there seems to be a third peak at 13.0Å in the solvent structure of the BOX simulation. From the solvent structure observed in these simulations the results of the simulation using the HAR boundary method differ from those of the SBMD and BOX simulations.

Fig. 11: Average Structure of Solvent Oxygens Around Ca^{+2} (Density Normalized Pair Correlation Function)



Molecular Dynamics and Peptide Flexibility

The molecular dynamics results of all three simulations are remarkably similar. The structure of the cyclic hexapeptide is quite rigid and the Ca^{2+} is tightly bound to the peptide complex. These structural properties reduce the effects of perturbations from the boundary constraints applied (to the buffer region) on the dynamics of the peptide-ion complex (in the reaction zone). The small time scale of these simulations (0.1ns) further restricts any effects of the three boundary methods on the dynamics of this system.

Figure 6 shows the distribution of peptide conformations in the last 20ps interval around the average peptide structure (calculated from this interval) for each of the three simulations. All the distributions center around 0.2-0.3Å. While the flexibility of the peptide is similar in all three simulations, the SBMD simulation does not have the larger tail that is found in the HAR simulation. The SBMD distribution has a smaller variance which suggests that the more deformable boundary method (SBMD) results in less perturbation to the reaction zone (i.e. the peptide structure). The HAR simulation has the broadest distribution and the largest tail (population at larger rms differences) suggesting that this boundary method causes a greater perturbation of the peptide structure.

The mobilities (positional fluctuations) of selected parts of the structure, from the three simulations, are shown in Figure 12. The BOX simulation allows the greatest mobility to the system which is reflected in these figures. This is especially noticeable in the solvent (which is unconstrained in this boundary method).

The dynamic property of the Ca^{2+}-peptide ligand interactions are shown in Figure 13. The peptide carbonyl oxygens that coordinate the Ca^{2+} are tightly bound and exhibit small, high frequency fluctuations. The similarity of the dynamics of the Ca^{2+}-

Fig. 12: RMS Fluctuations of Selected Groups of Atoms

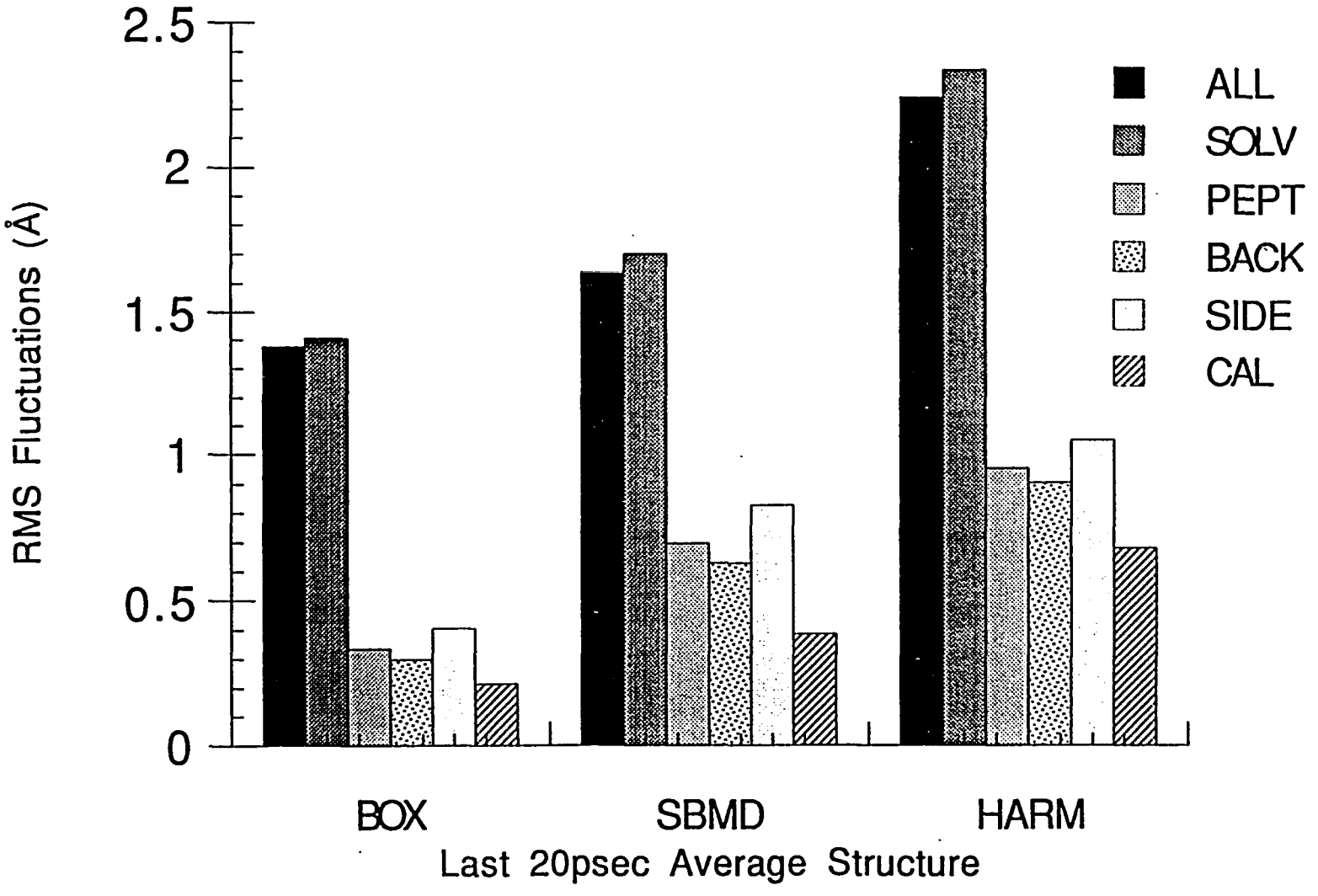


Fig 13a: BOX Time Series of Ca^{+2} -Peptide Ligand Distances

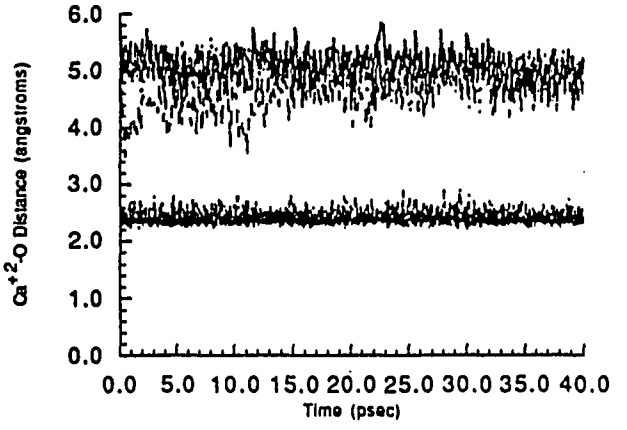


Fig. 13b: HARM Time Series of Ca^{+2}-Peptide Ligand Distances

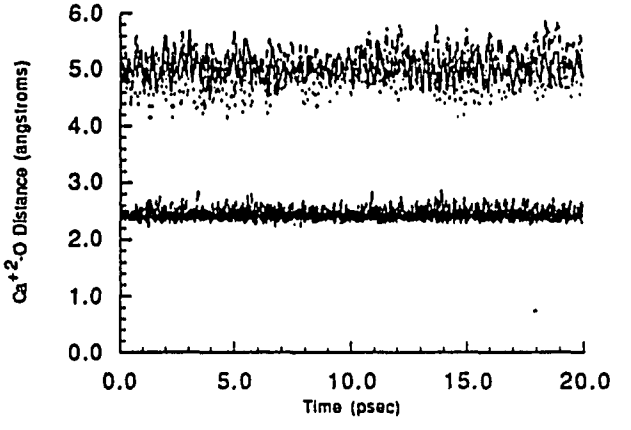


Fig 13c: SBMD Time Series of Ca^{+2}- Peptide Ligand Distances

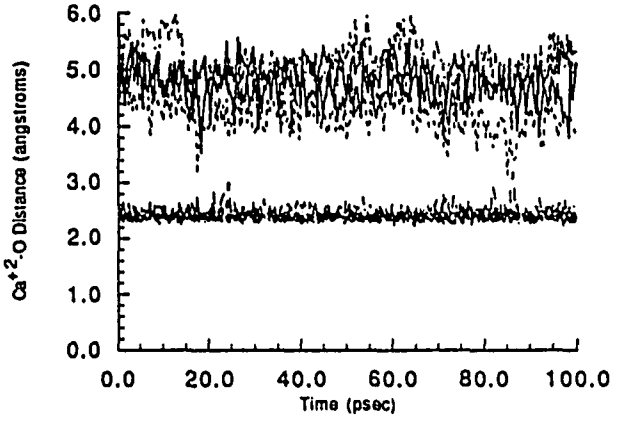


Fig. 14a: BOX Velocity Autocorrelation Function of Ca^{+2}

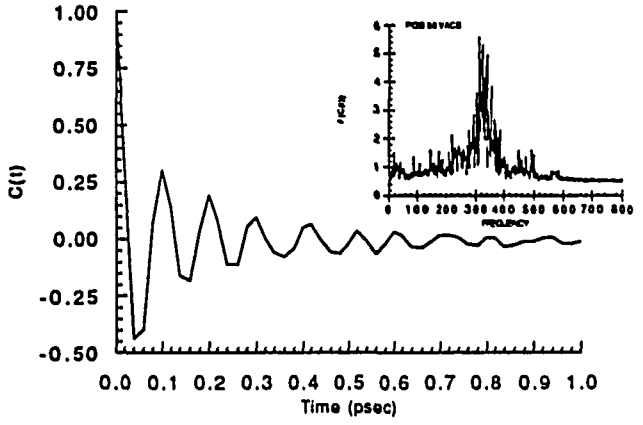


Fig. 14b: HARM Velocity Autocorrelation Function of Ca^{+2}

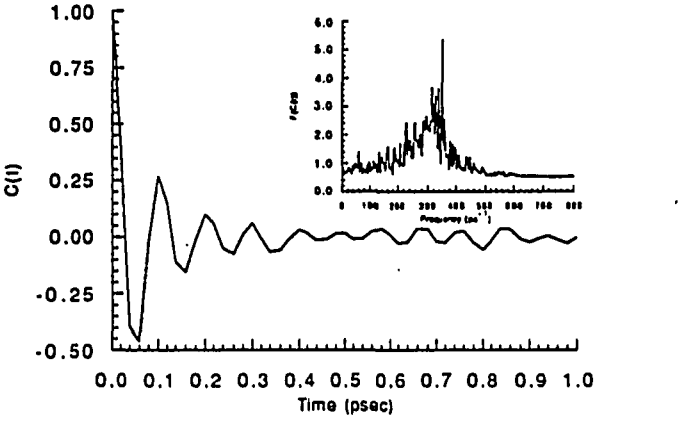
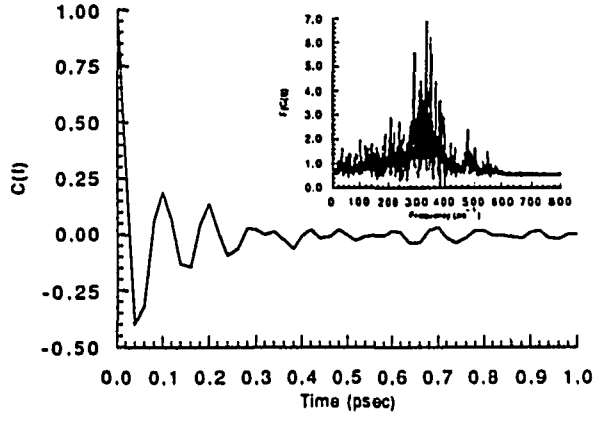


Fig. 14c: SBMD Velocity Autocorrelation function of Ca^{+2}



peptide coordination, among all three simulations, can be clearly seen in the velocity autocorrelation functions shown in Figure 14.

B. Effects of SBMD Boundary on Simulations of Solvated Erabutoxin

A comparison of molecular dynamics simulations of a protein, solvated with and without the stochastic boundary (SBMD) conditions, was performed. The results of these simulations provided an analysis of the effects of the SBMD boundary conditions on the structure and dynamics of an unconstrained protein. The purpose of this study was to provide evidence for the assumption that the effect of the SBMD boundary method on the reaction zone could be neglected in the analysis of simulations of larger macromolecular systems (i.e. solvated proteins). The unbounded solvated protein remained solvated throughout the entire simulation. While the solvent at the boundary (outside edge) of the system is affected by the absence of boundary conditions, it was assumed that the protein would experience little or no perturbations from these effects on the solvent and that the model represents a good approximation of a fully solvated protein. For these reasons the analysis of the results of this study focuses on the protein itself, rather than on the water structure. For comparison purposes, a simulation was performed of the protein *in vacuo*.

Solvation of Erabutoxin Xray Structure

The xray crystallographic structure of erabutoxin b (a snake venom) contains 111 crystal waters and was kindly provided by Dr. Low (Columbia University) [Bourne, 1985; Smith, 1988]. Construction of the starting structures in these simulations of solvated erabutoxin closely follows the procedure used in the study of PG3 (see above). The system was centered and a 30Å sphere of thermally equilibrated sphere of solvent was overlaid. Solvent overlapping the original system or greater than 28.96Å from the origin was removed. This system was the starting structure for both solvated

simulations. The structure of the simulation *in vacuo* was obtained by removal of all waters from the xray structure. Prior to performing molecular dynamics simulations these structures were briefly optimized with the CHARM empirical potential function in order to remove strain energy present. The optimization of the solvated simulation with the SBMD boundary conditions, called EBS, was performed after applying the boundary conditions, while the *in vacuo* simulation, called EB, and the unbounded solvated simulation, called EBU, were optimized without boundary conditions. The optimization procedure for the solvated simulations was 10 steps of conjugate gradient minimization of the waters only (protein fixed) followed by another 10 steps of minimization of all atoms (unconstrained). The procedure for the simulation *in vacuo* was 10 steps of unconstrained conjugate gradient minimization.

The EB and EBU simulations were first heated from 0°K to 300°K over 3ps (5°/20 steps) using the Verlet algorithm and then equilibrated. The EBS simulation started with equilibration at 300°K using Langevin dynamics and included a spherical buffer region of 3Å (all waters 25.96-28.96Å from the origin) with a buffer temperature scaling frequency of 100 steps. All three simulations used a non-bonded list update of 10 steps and were equilibrated for 10ps with an equilibration (temperature check/scaling) update every 10 steps. The SBMD boundary conditions were the same as described in the PG3 study except using a potential of mean force recalculated for a 28.96Å sphere. As in the PG3 study, this potential included only van der Waals interactions between solvent oxygens.

Equilibration of Protein Simulations

The total energy, potential energy and temperature plots from these simulations are shown in Figures 15-17. The plots of the solvated simulations show that they have equilibrated with kinetic energy fluctuations on the order of 1% of the total energy. These fluctuations are expected, from statistical mechanical fluctuation theory

Fig. 15a: EB In Vacuo MD of Erabutoxin b

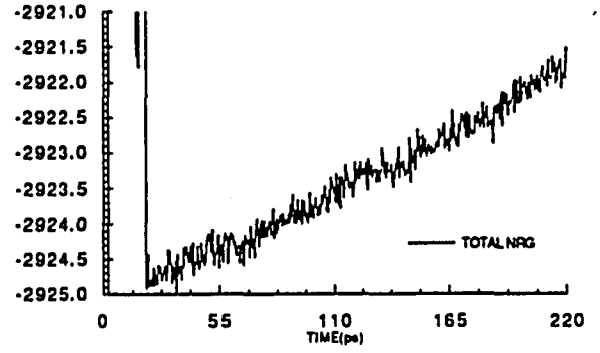


Fig. 15b: EBS Total Energy of Erabutoxin b SBMD

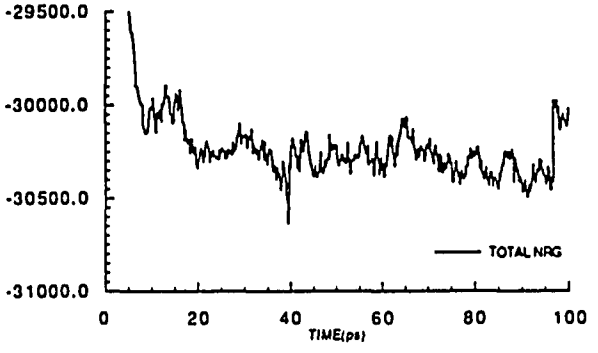


Fig 15c: EBU Unbounded, Solvated Erabutoxin b MD Total Energy

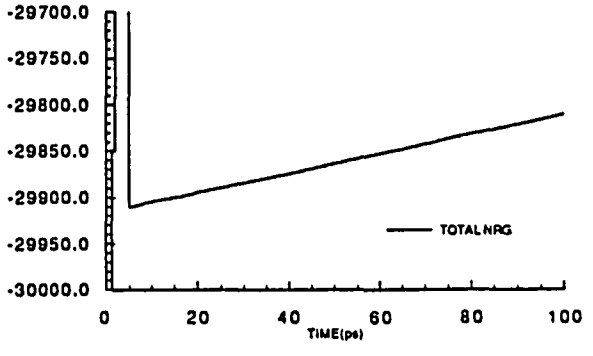


Fig. 16a: EB In Vacuo MD of Erabutoxin b

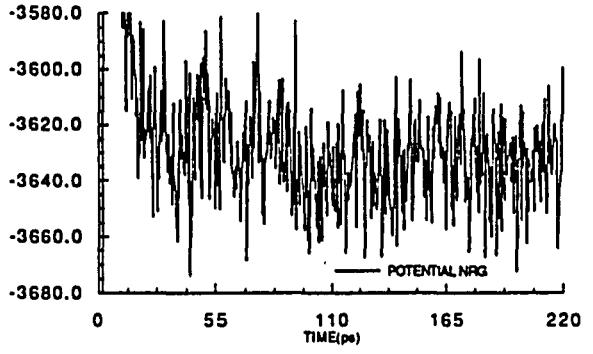


Fig. 16b: EBS Potential Energy of Erabutoxin b SBMD

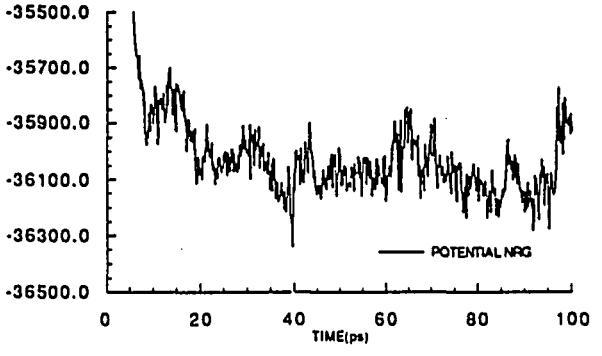


Fig. 16c: EBU Unbounded, Solvated Erabutoxin b MD Potential Energy

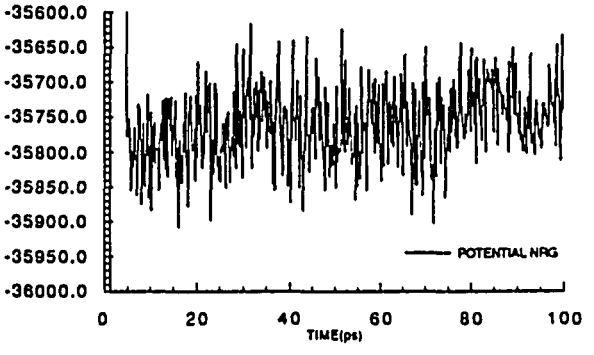


Fig. 17a: EB In Vacuo MD of Erabutoxin b

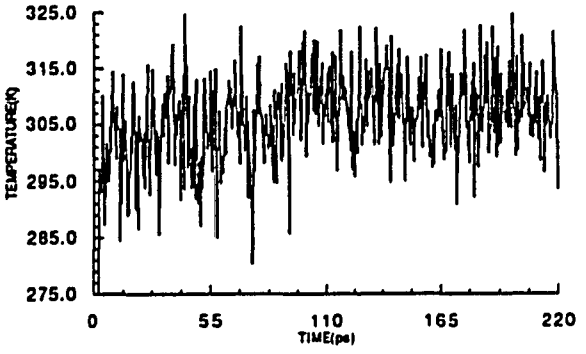


Fig. 17b: EBS Temperature of Erabutoxin b SBMD

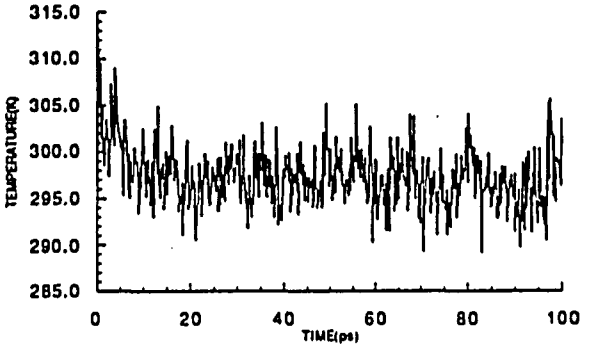


Fig. 17c: EBU Unbounded, Solvated Erabutoxin b MD Temperature

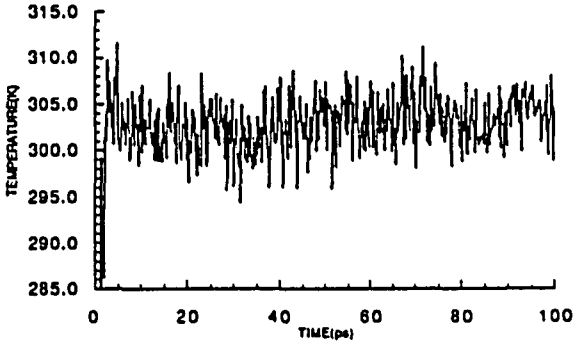
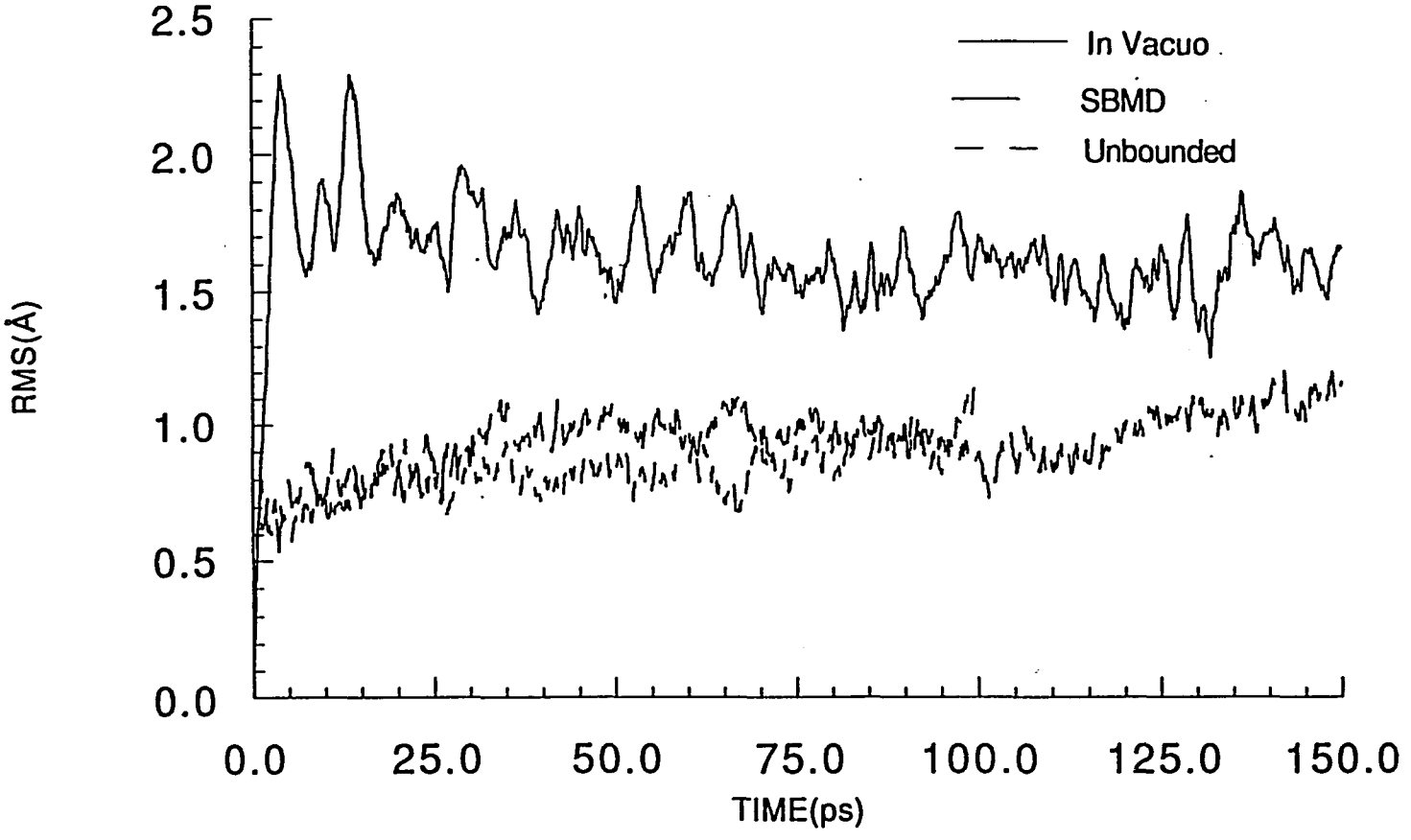


Fig. 18: RMS(CA) from X-ray of Erabutoxin b MD Trajectories



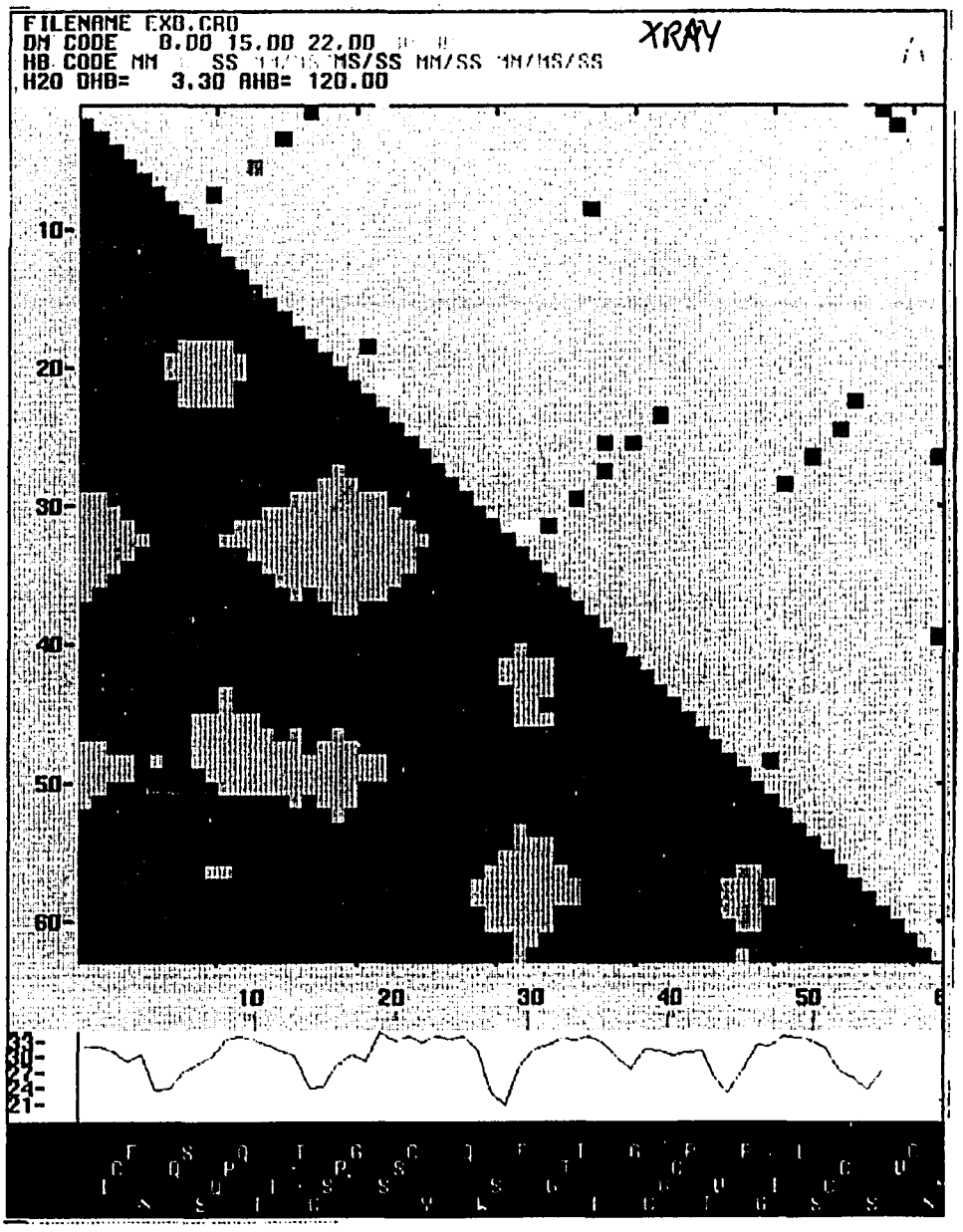


Figure 19a. DM/HBM/LDP plots of erabutoxin xray structure. DM is in bottom half of matrix, HBM is in top half of matrix, LDP is below matrix and primary sequence is at bottom of plot. Contours of DM (in angstroms) are 0-8(green), 8-15(light red), 15-22(dark red) and 22-30(orange). HB cutoffs are 3.3Å between heavy atoms and 120 degrees for A-H-D angle. HBM codes are MM(red), MS(yellow), SS(blue), MM/MS(orange), MS/SS(green), MM/SS(purple) and MM/MS/SS(brown). LDP is plotted in blue. Primary sequence residues are colored by polarity: non-polar(white), apolar(yellow), acidic(red) and basic(blue).

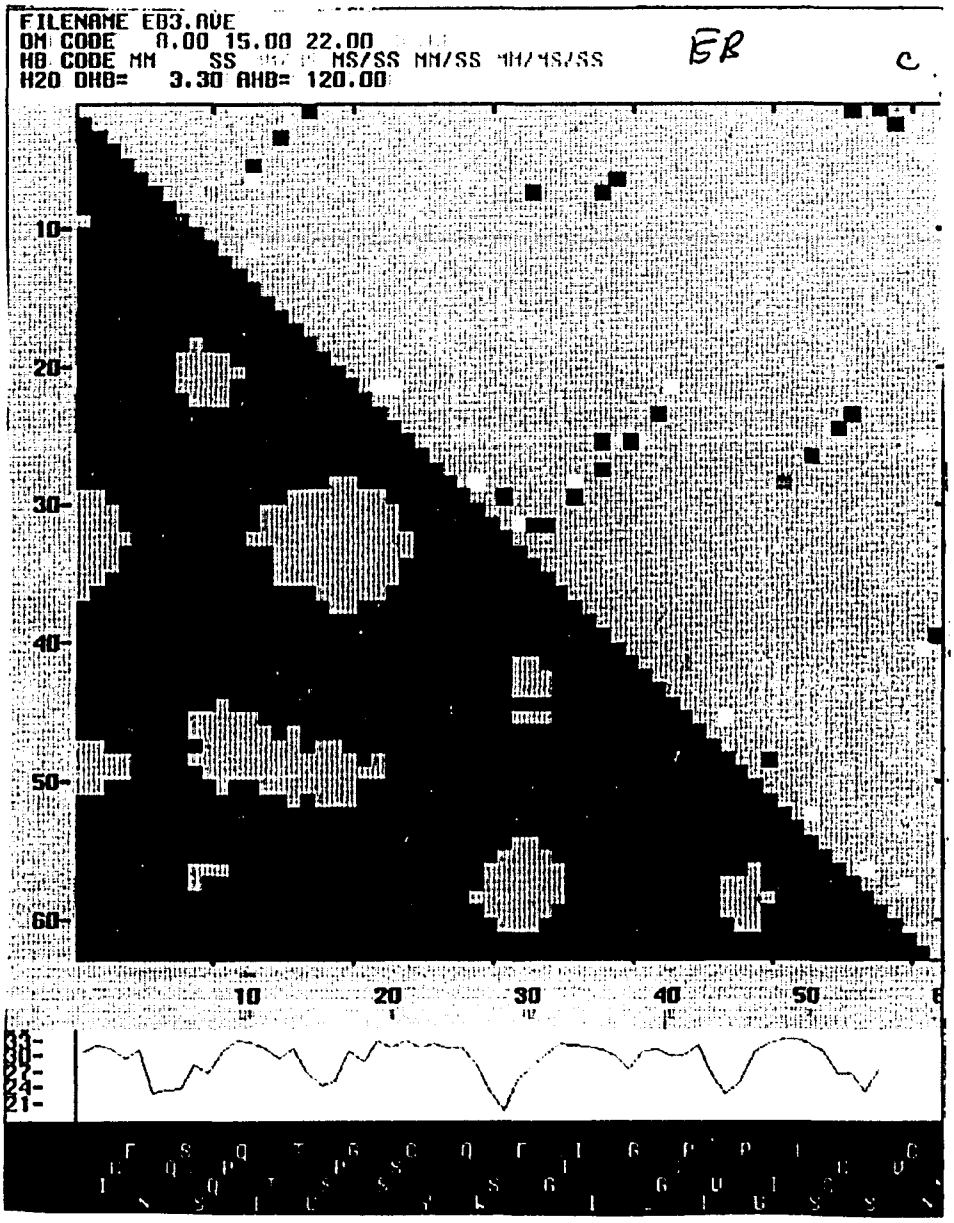


Figure 19b. DM/HBM/LDP plots of last 20ps average structure from EB (*in vacuo*) erabutoxin simulation. DM is in bottom half of matrix, HBM is in top half of matrix, LDP is below matrix and primary sequence is at bottom of plot. Contours of DM (in angstroms) are 0-8(green), 8-15(light red), 15-22(dark red) and 22-30(orange). HB cutoffs are 3.3Å between heavy atoms and 120 degrees for A-H-D angle. HBM codes are MM(red), MS(yellow), SS(blue), MM/MS(orange), MS/SS(green), MM/SS(purple) and MM/MS/SS(brown). LDP is plotted in blue. Primary sequence residues are colored by polarity: non-polar(white), apolar(yellow), acidic(red) and basic(blue).

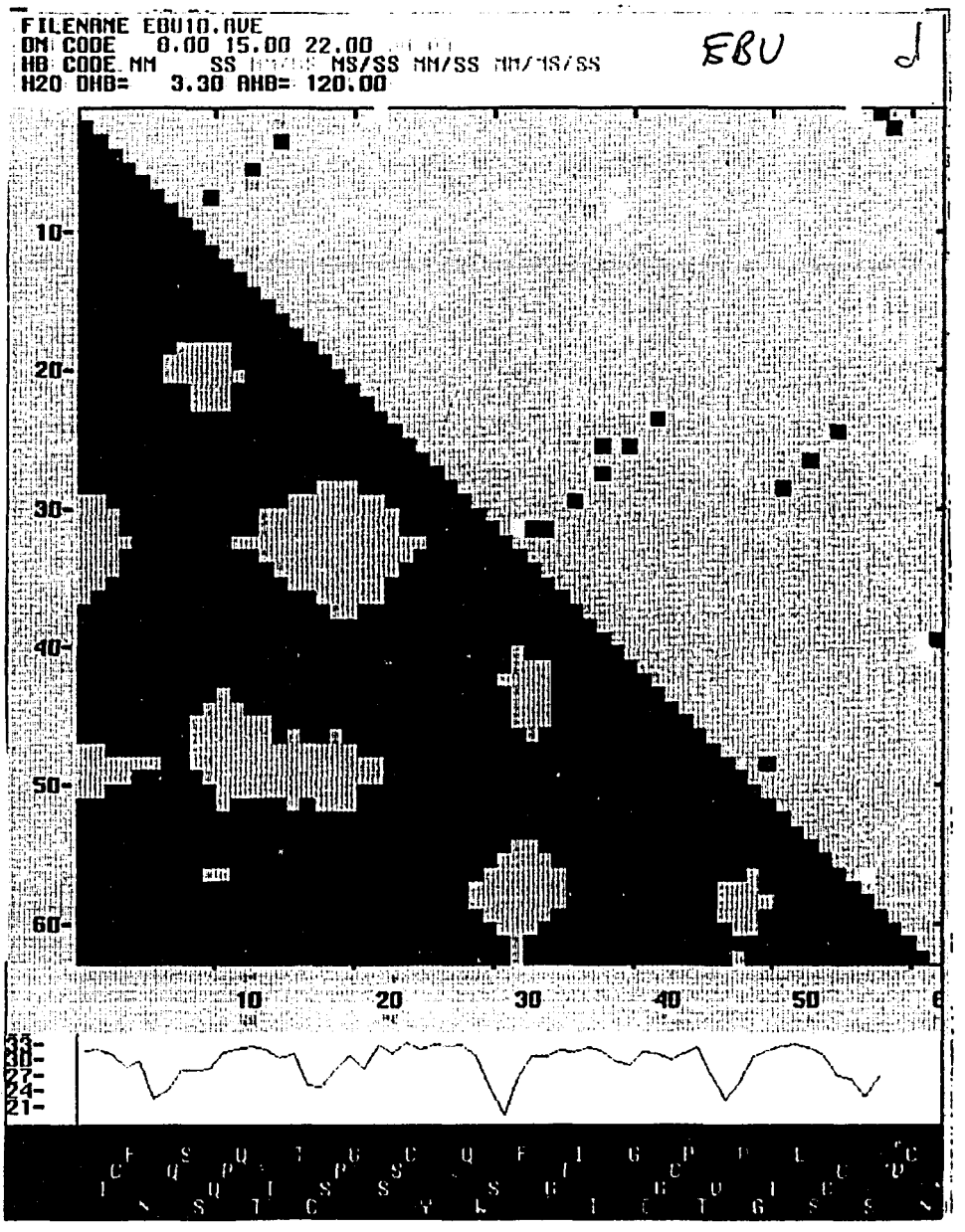


Figure 19c. DM/HBM/LDP plots of last 20ps average structure from EBU (unbounded solvated) erabutoxin simulation. DM is in bottom half of matrix, HBM is in top half of matrix, LDP is below matrix and primary sequence is at bottom of plot. Contours of DM (in angstroms) are 0-8(green), 8-15(light red), 15-22(dark red) and 22-30(orange). HB cutoffs are 3.3Å between heavy atoms and 120 degrees for A-H-D angle. HBM codes are MM(red), MS(yellow), SS(blue), MM/MS(orange), MS/SS(green), MM/SS(purple) and MM/MS/SS(brown). LDP is plotted in blue. Primary sequence residues are colored by polarity: non-polar(white), apolar(yellow), acidic(red) and basic(blue).

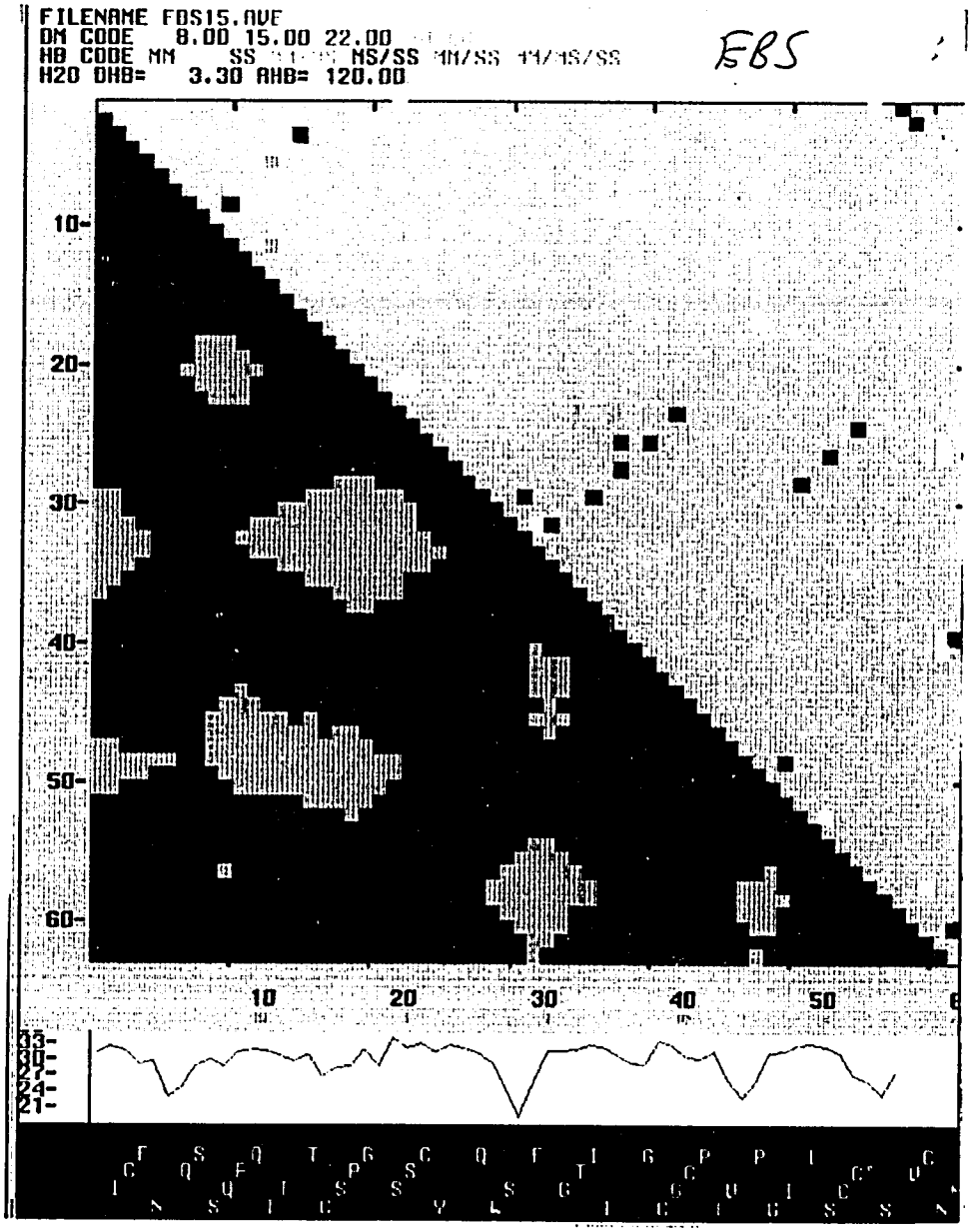


Figure 19d. DM/HBM/LDP plots of last 20ps average structure from EBS (SBMD) erabutoxin simulation. DM is in bottom half of matrix, HBM is in top half of matrix, LDP is below matrix and primary sequence is at bottom of plot. Contours of DM (in angstroms) are 0-8(green), 8-15(light red), 15-22(dark red) and 22-30(orange). HB cutoffs are 3.3Å between heavy atoms and 120 degrees for A-H-D angle. HBM codes are MM(red), MS(yellow), SS(blue), MM/MS(orange), MS/SS(green), MM/SS(purple) and MM/MS/SS(brown). LDP is plotted in blue. Primary sequence residues are colored by polarity: non-polar(white), apolar(yellow), acidic(red) and basic(blue).

[McQuarrie, 1988], for systems containing ca. 10,000 atoms. The drift in total energy and temperature in each simulation is on the order of 0.1%. The time point at which the simulations appeared to equilibrate is 40ps in EBU (unbounded solvated), 40ps in EBS (SBMD solvated) and 120ps in EB (in vacuo).

Figure 18 shows the rms difference in alpha carbon positions of the xray structure from the coordinate trajectories of the simulations. The change in the protein structures resulting from all three simulations (from the xray structure) appear to be stable at ca. 1.0Å rms. Figure 19 shows the HBM, DM and LDP plots of the xray and final average structures from the three simulations. There is very little loss of secondary structure and no large tertiary structural changes in the three simulations. The DLDP plots of the three simulations structures from the xray, shown in Figure 20, clearly show that the structural changes in the protein backbone resulting from the different simulations are actually small differences distributed throughout the structure and not large differences localized to small areas. In conclusion, while there are small measurable differences resulting from each simulation, the simulations have equilibrated and the xray conformation is well maintained.

Effect of Stochastic Boundary Conditions

Protein Structure, Protein Dynamics, Solvent Structure

Table 4 shows the rms difference in position of alpha carbons among the xray structure, the minimized starting structures and the final average structures (from the last 20ps of the simulations). The final average structure from the in vacuo simulation has an rms difference of greater than 1.4Å to the xray structure and the final average structure from both the solvated simulations. Table 4 shows an rms of 0.9Å between the final average structures from the two solvated simulations, indicating that the protein structures resulting from these two simulations are not significantly different. Both

Table 4. Alpha carbon rms positional differences from the among the xray structure, the minimized starting structures and the final average structures from the erabutoxin simulations.

	XRAY	EBMIN	EBAVE	EBUMIN	EBUAVE	EBSMIN	EBSAVE
XRAY	0.00						
EBMIN	0.09	0.00					
EBAVE	1.63	1.64	0.00				
EBUMIN	0.13	0.06	1.64	0.00			
EBUAVE	0.76	0.76	1.37	0.77	0.00		
EBSMIN	0.13	0.06	1.64	0.00	0.77	0.00	
EBSAVE	1.06	1.06	1.71	1.07	0.88	1.07	0.00

XRAY = Xray structure of erabutoxin b
EBMIN = Minimized starting structure of *in vacuo* simulation
EBAVE = Final average structure from *in vacuo* simulation
EBUMIN = Minimized starting structure of unbounded solvated simulation
EBUAVE = Final average structure from unbounded solvated simulation
EBSMIN = Minimized starting structure of SBMD solvated simulation
EBSAVE = Final average structure from SBMD solvated simulation

structures are are also similar to the xray structure (rms differences of 0.8Å or the unbounded and 1.1Å for the SBMD solvated structures).

The DLDP (difference plot) of the final average structures of the two solvated simulations show no significant difference in secondary structure and the DDM (difference plot) reveals the large similarity in their tertiary structures. The HBM plots in Figure 19 also show that the EBS simulation resulted in a protein structure similar to the EBU simulation.

The DLDP plot in Figure 21 shows the difference of the two solvated structures from the *in vacuo* structure. This plot shows the effect of solvent in these simulations on the secondary structure of erabutoxin. Notice that the effect is virtually identical for both simulations. Thus, when compared to the xray structure, modelling the solvated erabutoxin with the SBMD boundary method appears to have an identical effect as using unbounded solvent.

Fig. 20a: Difference LDPs from X-ray of Last 20ps Average Structures of Erabutoxin b Simulations

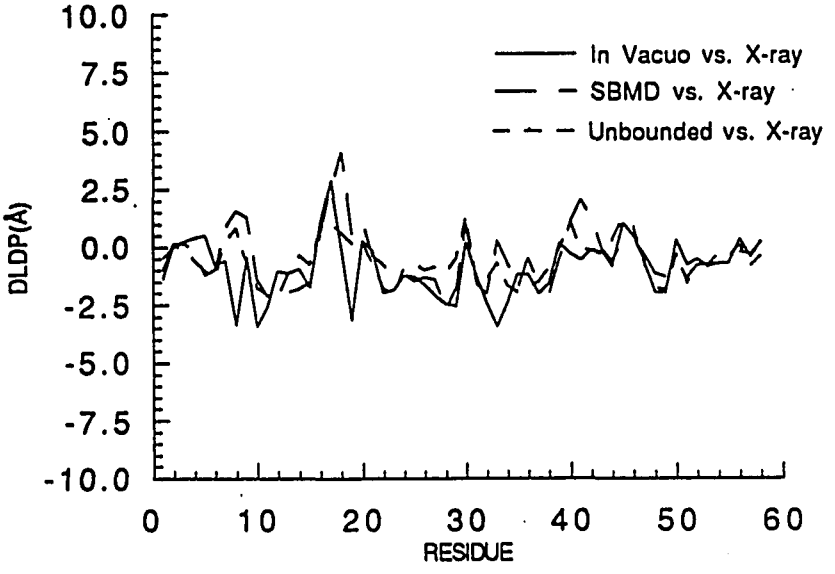


Fig. 20b: Difference CAT from X-ray of Last 20ps Average Structures from Erabutoxin b Simulations

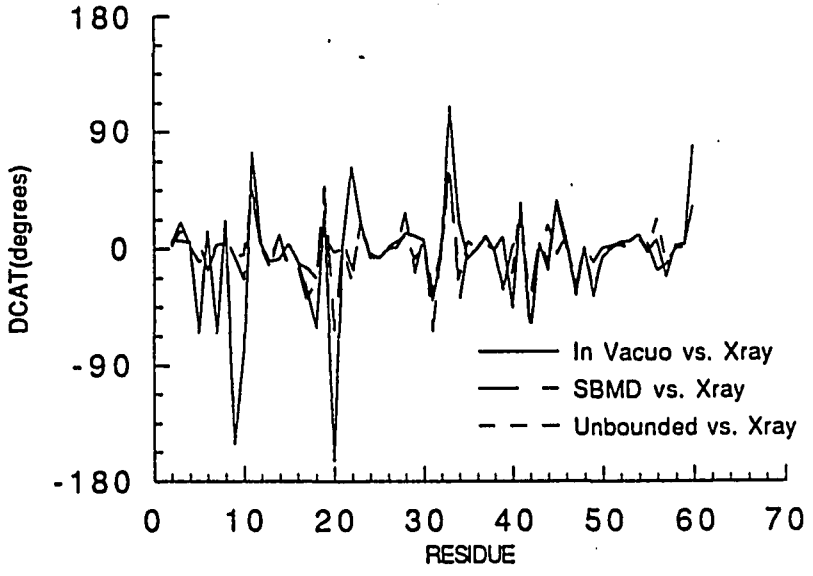


Fig. 21a: Difference LDPs of Last 20ps Average Structures from Solvated vs. In Vacuo Erabutoxin b Simulations

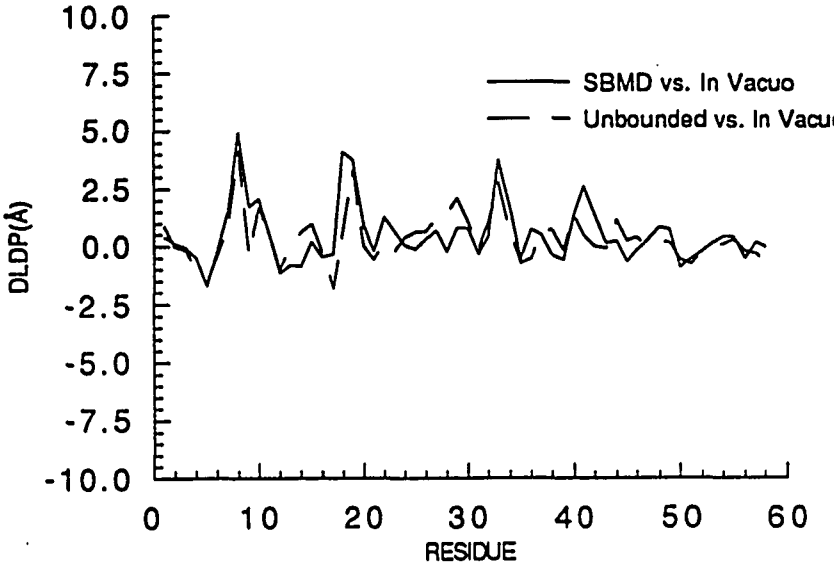
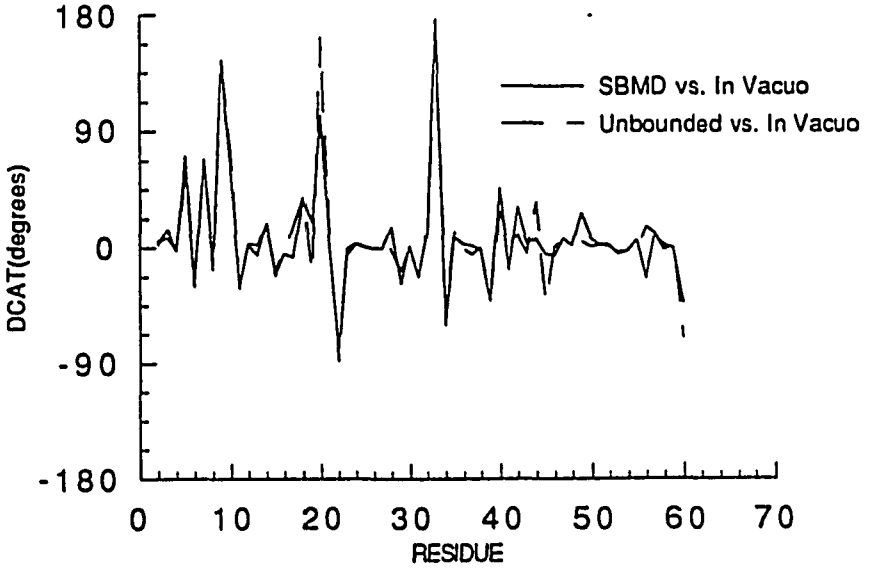


Fig 21b: Difference CAT of Last 20ps Average Structures of Solvated from In Vacuo Erabutoxin b Simulations



The B-value plots of backbone and sidechain mobility of erabutoxin from the xray data is obtained from Low, et. al. (Figure 2, 1986) while the B-values calculated from the atomic mobilities obtained from the coordinate trajectories of the EBS and EBU simulations of the solvated erabutoxin b are displayed in Figure 22. The plots from both simulations are very similar and apart from one area (residues 30-35) are in close agreement with the xray data. From the results of these two solvated simulations it is shown that whether or not the SBMD boundary conditions are imposed there is no significant difference in the molecular dynamics of the solvated protein.

A model for the mechanism of erabutoxin b binding to the acetylcholine receptor has been proposed by Low et. al. [Low, 1986, 1987]. Results of the present study have implications for this hypothesis and are discussed in the Conclusions section. The difference between the xray B-value data and the results of the solvated simulations (at residues 45-55) relate to the proposed mechanism of binding and may suggest that the lower mobility of this segment observed in the xray data result from crystal packing effects or crystallization conditions.

The results of this study, however, comparing the two methods of simulating solvated erabutoxin do support the use of the SBMD boundary method in simulations of large macromolecular structures. The radial distribution functions for solvent from the origin, from the last 20ps of the two solvated simulations, are shown in Figure 23. These data do not indicate any large differences in solvent structure between the two simulations.

Fig. 22a: EB B Values from Last 20ps of Erabutoxin b In Vacuo Simulation

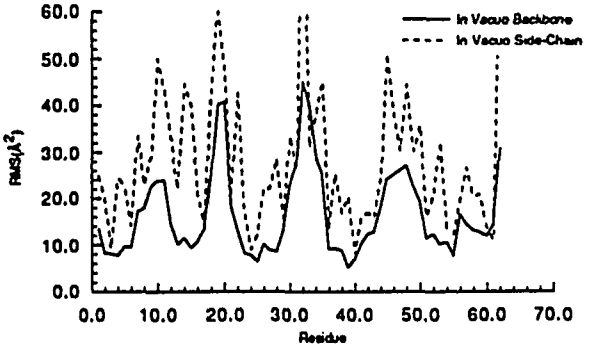


Fig. 22b: EBS B Values from Last 20ps of Erabutoxin b SBMD Simulation

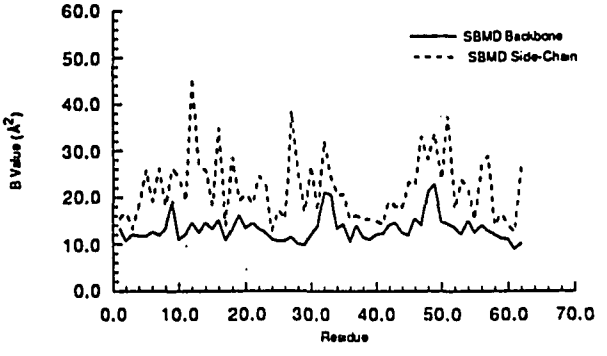


Fig. 22c: B Values from Last 20ps of Erabutoxin b Unbounded Simulation

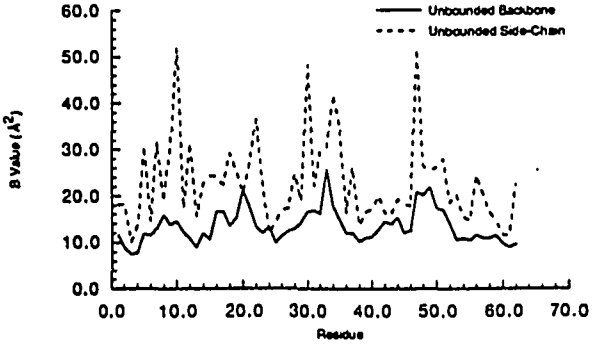
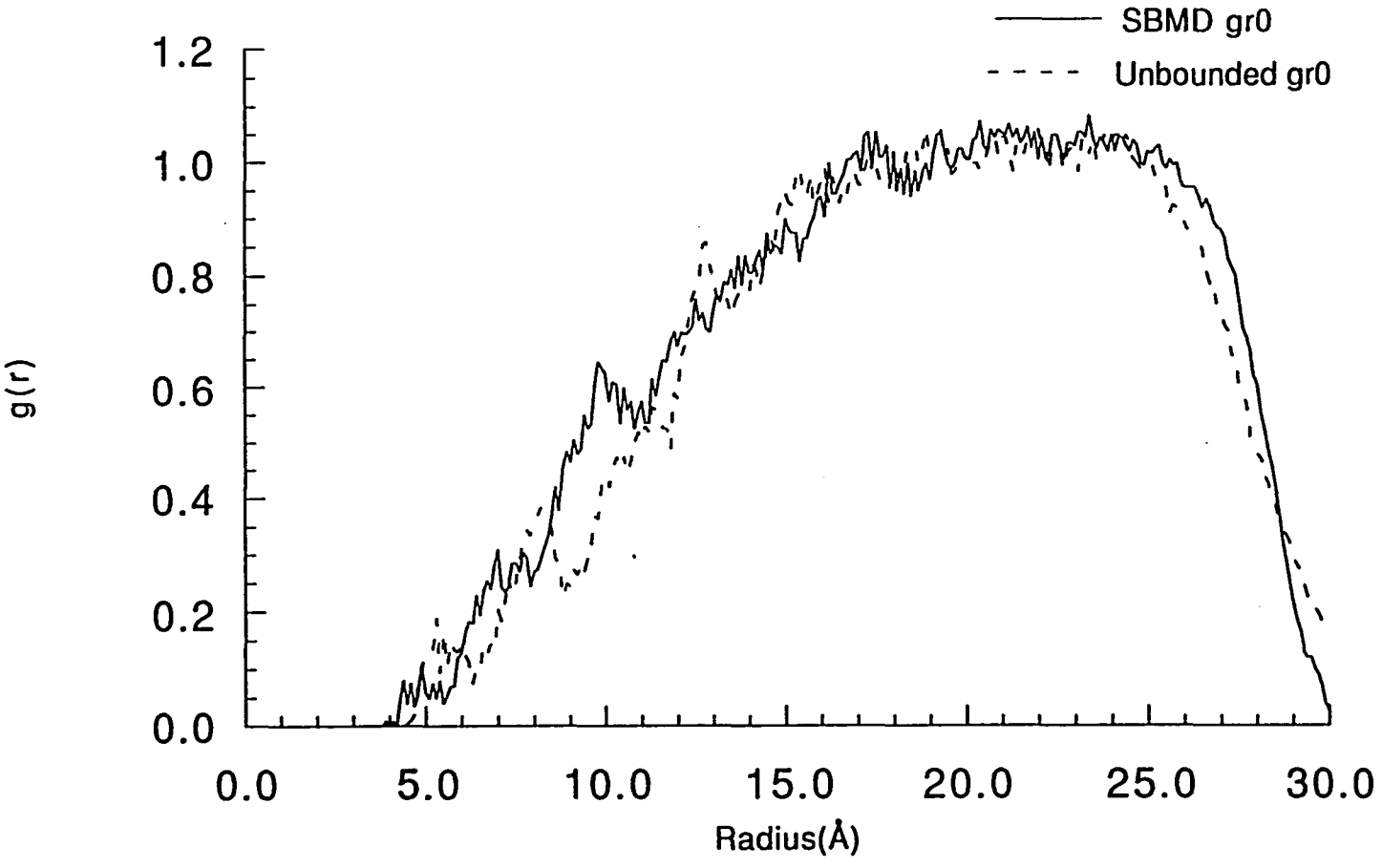


Fig. 23: Water Structure (g(OO)) of Last 20ps of Solvated Erabutoxin b Simulations



C. Comparison of *In Vacuo* and SBMD Simulations of Subtilisin BPN'

Before studying of the role of serine 163 in the mechanism of Ca^{2+} binding to site B in subtilisin, the effect of solvent on the structure and dynamics of the Ca^{2+} binding site was explored by a comparison of a solvated simulation of this structure with an *in vacuo* simulation. Both simulations used a high resolution xray structure of subtilisin BPN', kindly provided by Dr. Krauss [Nat. Bureau Standards, Washington, D.C.] as the starting conformation. The results of the simulation of the solvated Ca^{2+} binding site, using the SBMD boundary method, provided the starting structure for the subsequent study on the role of serine 163. This study explored the role of solvent in the structure and dynamics of Ca^{2+} site by comparing the structures and molecular dynamics that were obtained from these two simulations.

Construction of Solvated Xray Calcium Ion Binding Site

The simulation of the solvated subtilisin Ca^{2+} binding site B, called SBT30, used the SBMD boundary method described in the previous studies [Brooks, 1983a]. The system was solvated by overlaying a 30.0Å sphere of thermalized waters centered on the Ca^{2+} in site B. Solvent overlapping the original system or greater than 29.71Å from the origin was removed. The SBT30 structure consisted of subtilisin and 2512 water molecules. The *in vacuo* simulation, called SBTV, consisted of the xray structure (protein and 218 crystal waters) simulated with Verlet molecular dynamics and a distance dependent dielectric. Prior to starting the *in vacuo* simulation strain energy in the xray structure was removed through an optimization procedure in which the protein atoms were constrained with harmonic potential set to the reference (i.e. xray) position. These constraints were reduced gradually over multiple conjugate gradient minimizations after which there was a final unconstrained minimization.

In the solvated simulation of the Ca^{2+} binding site, as opposed to the previous simulations of solvated erabutoxin and PG3, the protein extended well beyond the reaction zone and into the reservior. Thus it was necessary to use reduced dynamics which included boundary constraints on protein atoms (outside the reaction zone). Here for the first time it was necessary to constrain protein atoms in the buffer zone (defined as 26.71-29.71Å from the Ca^{2+} in site B); protein atoms in the buffer were restrained with a harmonic restraining potential proportional to their distance from the Ca^{2+} using empirically determined parameters [Brooks, 1985]. The exact form of the potential is shown in equation 10 (see Methods, section B). Langevin dynamics were applied to the protein atoms in the buffer using a frictional constant of 62.5/psec [Brooks, 1985]. Protein residues and crystal waters in the reservoir (>29.96Å from the Ca^{2+} in site B) were removed; the reservoir residues removed were 1, 2, 37 to 41, 55, 76, to 80 and 211 to 214. The stochastic potential of mean force modelling the bulk solvent, as in the previous simulations using the SBMD boundary method, applied only to van der Waals interactions between solvent oxygen atoms. For the solvated simulation the starting structure was optimized, to reduce strain energy, by conjugate gradient minimizations first with the protein constrained and then an unconstrained minimization. The molecular dynamics simulation of the solvated structure began with an equilibration of the solvent. During the first 20ps of the molecular dynamics simulation the protein atoms were constrained (fixed) and only the solvent was equilibrated. The unconstrained, solvated simulation began after this period of solvent equilibration.

Equilibration of Protein Structures *In Vacuo* and in Solvent

In the *in vacuo* simulation the system (i.e. protein and crystal waters) was unconstrained. A total of 82 of the 218 crystal waters dissociated from the protein

during equilibration. The other 136 water molecules remained with the protein during the production phase of the simulation.

Energy plots for the two simulations are shown in Figures 24 and 25. Temperature fluctuations and total energy drift are within acceptable limits. The *in vacuo* simulation equilibrated at ca. 50ps, while in the solvated simulation the temperature plot indicates changes occurring after several hundred picoseconds.

Rms alpha carbon positional differences, from the xray structure, observed in the trajectories of these simulations are shown in Figure 26. While the *in vacuo* simulation has a final difference from the xray structure of ca. 1.6Å there is an oscillatory behavior to the dynamics of the protein around this value. The solvated simulation results in an ca. 2.0Å difference from the xray structure with no such oscillations observed in the behavior of this structure. The major conformational change in the protein is in the tertiary structural orientation of the N-terminal domain (residues 1-100) in the buffer region and the two carboxy domains (residues 101 to 275) in the reaction zone. The discussion below suggests that these changes occurred as a result of the difference in the effects of solvent on the specific domains of the protein. Changes in the coordination of the Ca^{2+}, discussed below, occurred between 230-300ps of the trajectory, and the drift in conformation of the protein, measured from the xray structure, is small after 350ps. While further conformational changes in the solvated protein structure cannot be ruled out, this analysis suggests that the structural change after 400ps between the reaction zone and buffer (i.e., the interdomain tertiary structure) is due to dynamics in which the protein is sampling local conformational space.

Fig. 24a: SBT SBMD Total Energy vs. Time

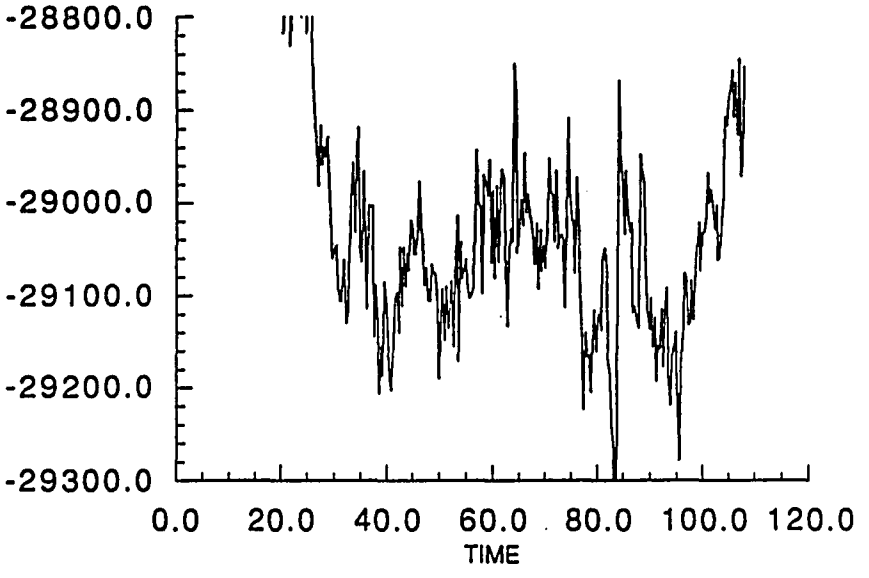


Fig. 24b: SBT Total Energy vs. Time for In Vacuo Simulation

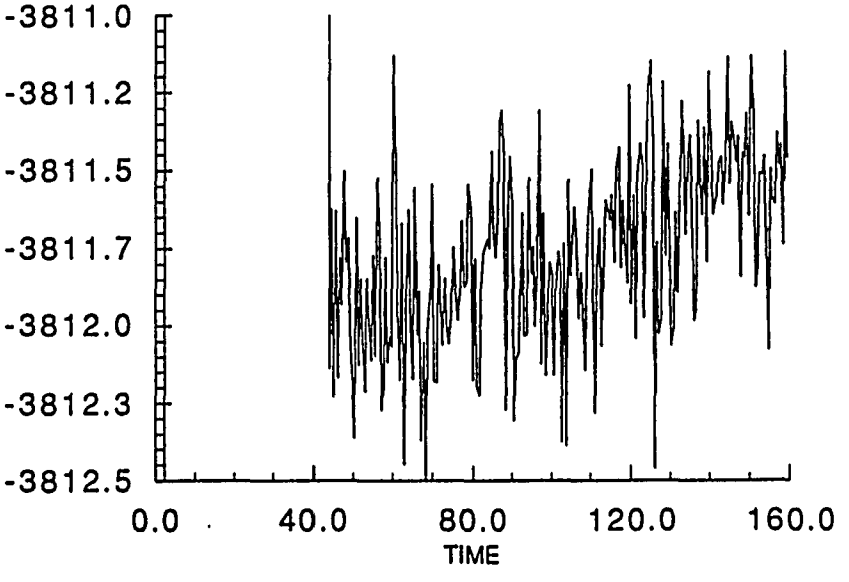


Fig. 25a: SBT SBMD Temperature vs. Time

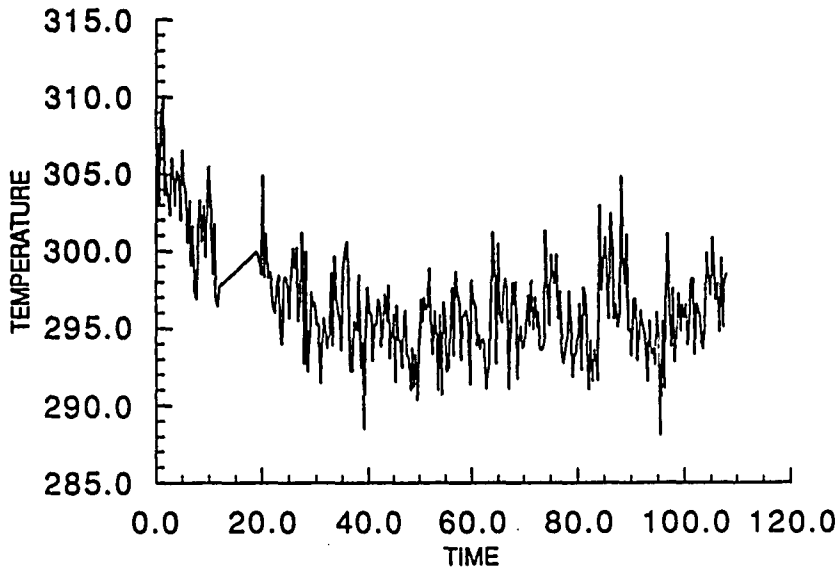


Fig. 25b: SBT Temperature vs. Time for In Vacuo Simulation

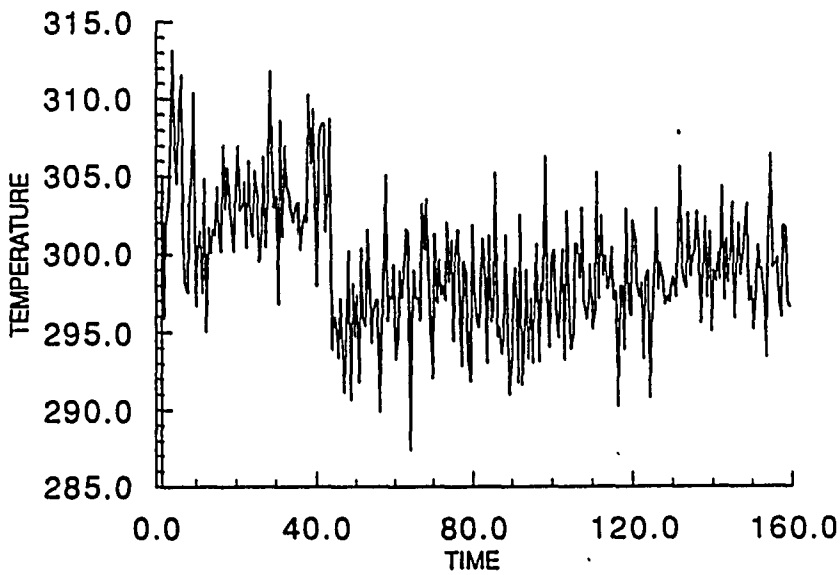


Fig. 26a: SBT30 RMS(CA) vs. Minimized Starting Structure of Reaction Zone of Solvated SBMD Subtilisin BPN' Simulation

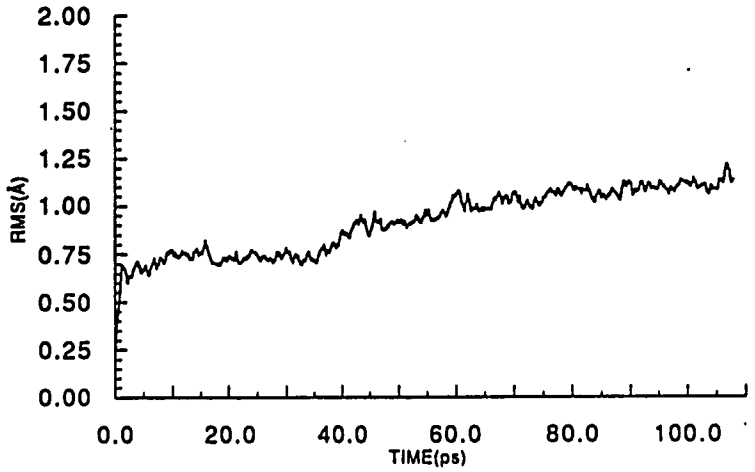
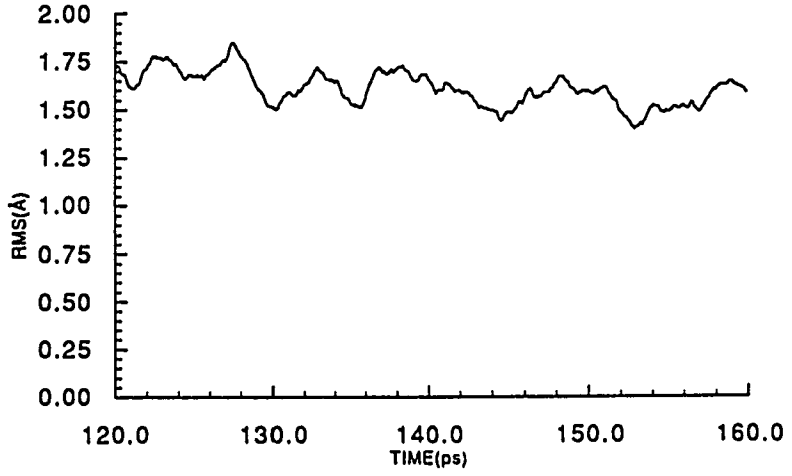


Fig. 26b: SBTV RMS(CA) from X-Ray of In Vacuo Subtilisin BPN' Simulation



Comparison of Protein Structures *In Vacuo* and in Solvent

Structure of Ca^{2+} Binding Site B

The alpha carbon rms positional difference between the xray structure and the final average structure of the *in vacuo* simulation was 1.40Å. The 420-440ps average structure is representative of the production phase of this simulation (i.e. after 400ps). It is used in the analysis here since the structure at 440ps was the starting structure used in the following study. The rms positional difference of alpha carbons in the reaction zone of the solvated structure was 1.81Å to the xray structure and 1.93Å to the *in vacuo* structure.

While the overall structural change of the solvated protein is ca. 2.0Å, shown in Figure 26, analysis of the conformational changes that occurred reveals that this difference is largely the result of a separation between the N-terminal domain (whose structure is constrained) and the two domains that are in the reaction zone. The LDP and DM plots, Figures 27 and 28, show that the conformational change is not due to an isolated perturbation in the solvated protein and that the solvated structure maintains the overall tertiary and secondary elements of the xray conformation. The DDM plots in Figure 29 show that the first domain (ca. residues 1-100) in the average solvated structure, many of whose residues are in the buffer region, differs very little from the xray structure, but has separated by ca. 2Å rms from the two domains in the reaction zone (ca. residues 101-275). While the N-terminal domain is constrained in the solvated simulation, this domain does undergo conformational changes in the *in vacuo* simulation. The two carboxy domains in the reaction zone of the solvated protein interact freely with the explicit solvent molecules in that simulation while the domains in the buffer are constrained. The change in tertiary structure between the N-terminal domain in the buffer and the reaction zone in the solvated simulation is probably an artifact of the boundary conditions imposed. However this minor

Fig. 27a: LDP of Xray Structure of Subtilisin

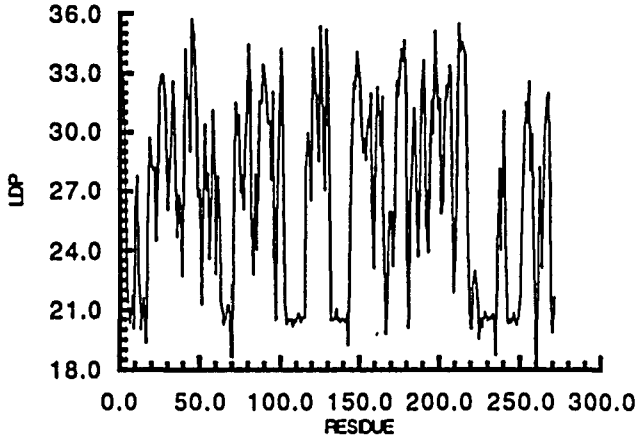


Fig. 27b: LDP from In Vacuo Simulation of Subtilisin

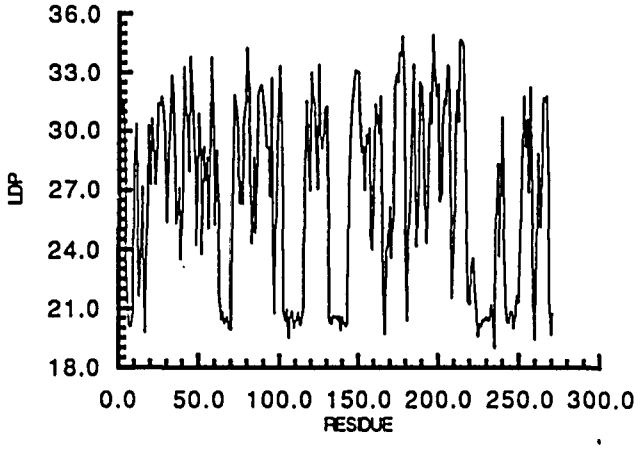
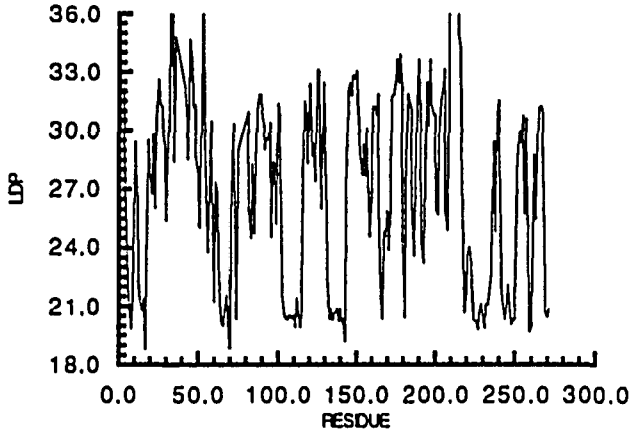


Fig. 27c: LDP from Solvated Simulation of Subtilisin



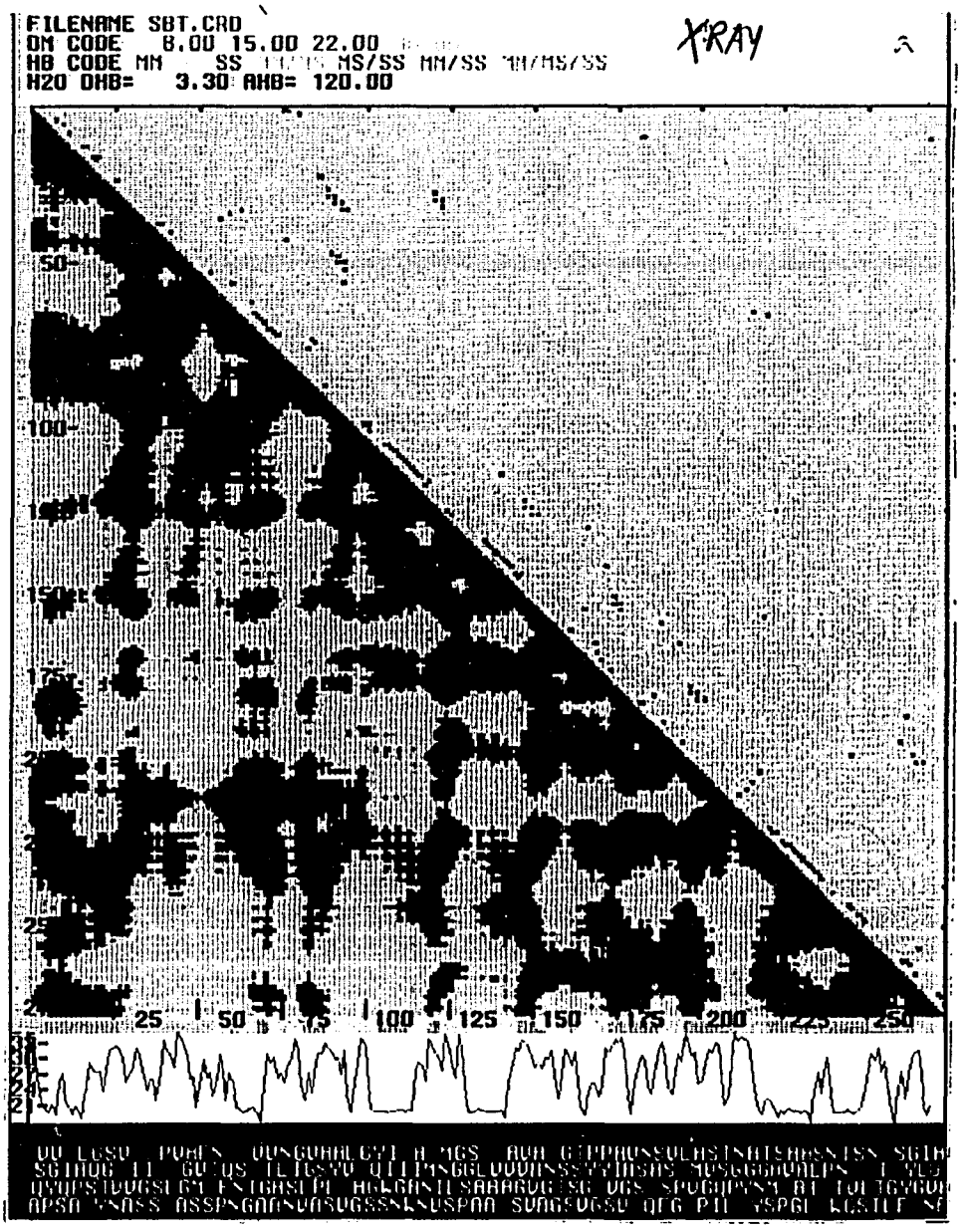


Figure 28a. DM/HBM/LDP plots of subtilisin xray structure. DM is in bottom half of matrix, HBM is in top half of matrix, LDP is below matrix and primary sequence is at bottom of plot. Contours of DM (in angstroms) are 0-8(green), 8-15(light red), 15-22(dark red) and 22-30(orange). HB cutoffs are 3.3Å between heavy atoms and 120 degrees for A-H-D angle. HBM codes are MM(red), MS(yellow), SS(blue), MM/MS(orange), MS/SS(green), MM/SS(purple) and MM/MS/SS(brown). LDP is plotted in blue. Primary sequence residues are colored by polarity: non-polar(white), apolar(yellow), acidic(red) and basic(blue).

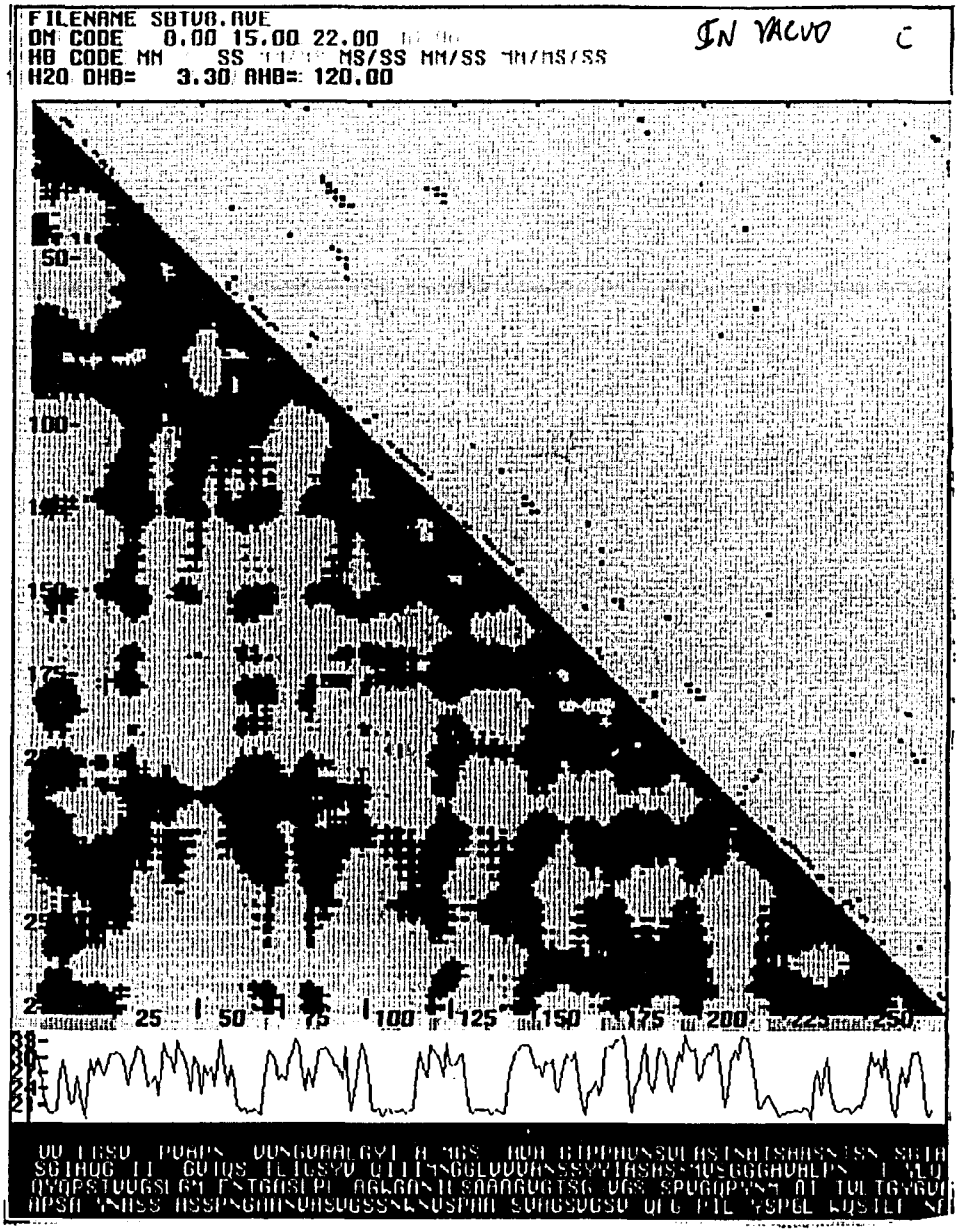


Figure 28b. DM/HBM/LDP plots of last 20ps average structure from SBTV (*in vacuo*) subtilisin simulation. DM is in bottom half of matrix, HBM is in top half of matrix, LDP is below matrix and primary sequence is at bottom of plot. Contours of DM are 0-8Å(green), 8-15Å(light red), 15-22Å(dark red) and 22-30Å(orange). HB cutoffs are 3.3Å between heavy atoms and 120 degrees for A-H-D angle. HBM colors are MM(red), MS(yellow), SS(blue), MM/MS(orange), MS/SS(green), MM/SS(purple) and MM/MS/SS(brown). LDP is plotted in blue. Primary sequence residues are colored by polarity: non-polar(white), apolar(yellow), acidic(red) and basic(blue).

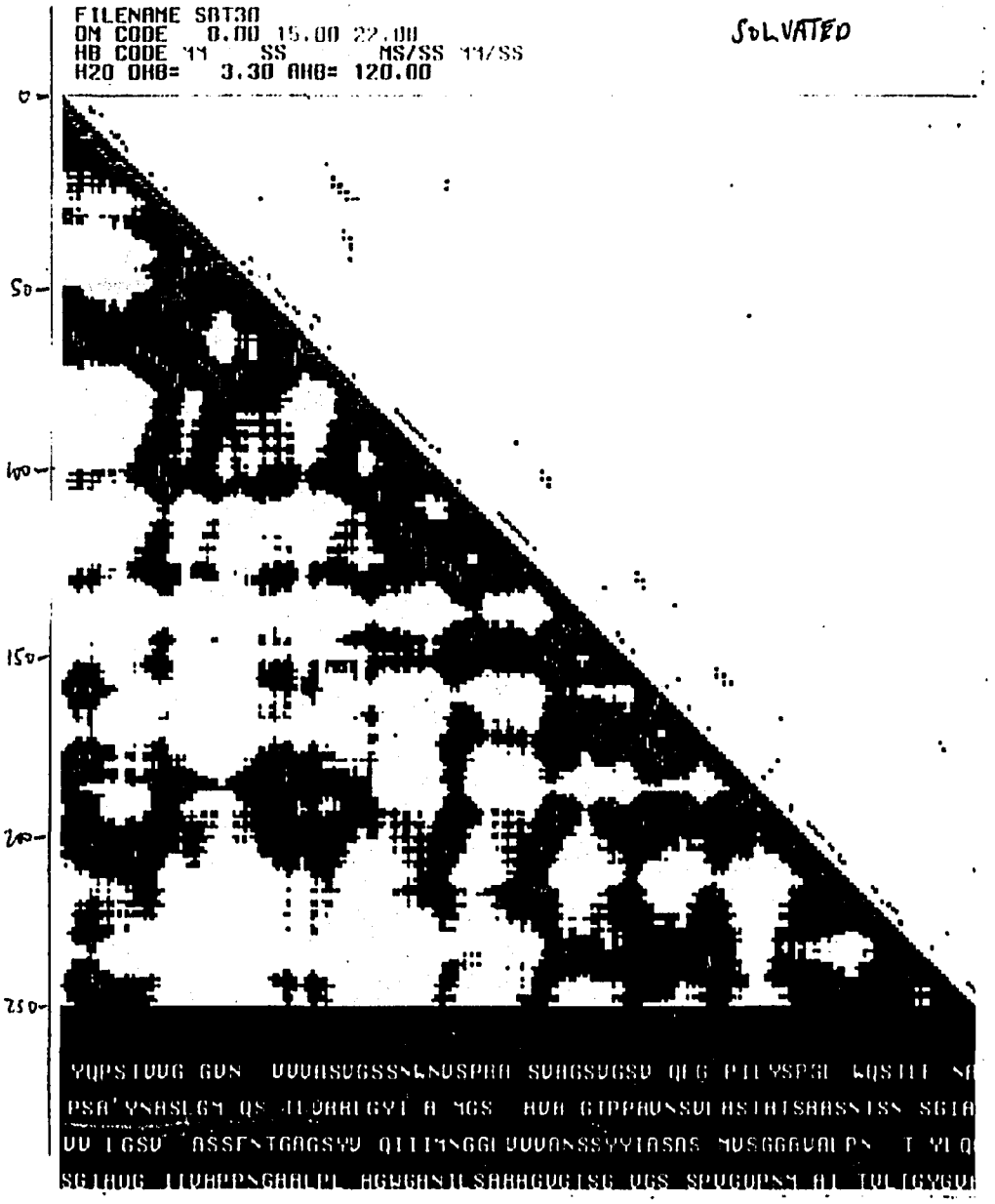


Figure 28c. DM/HBM plots of last 20ps average structure from SBT30 (SBMD) subtilisin reduced dynamics simulation. There are only 258 residues in this structure; primary sequence numbers from native structure are shown in parentheses. DM is in bottom half of matrix, HBM is in top half of matrix and primary sequence is at bottom of plot. Contours of DM are 0-8Å(green), 8-15Å(light red), 15-22Å(dark red) and 22-30Å(orange). HB cutoffs are 3.3Å between heavy atoms and 120 degrees for A-H-D angle. HBM colors are MM(red), MS(yellow), SS(blue), MM/MS(orange), MS/SS(green), MM/SS(purple) and MM/MS/SS(brown). Primary sequence residues are colored by polarity: non-polar(white), apolar(yellow), acidic(red) and basic(blue).

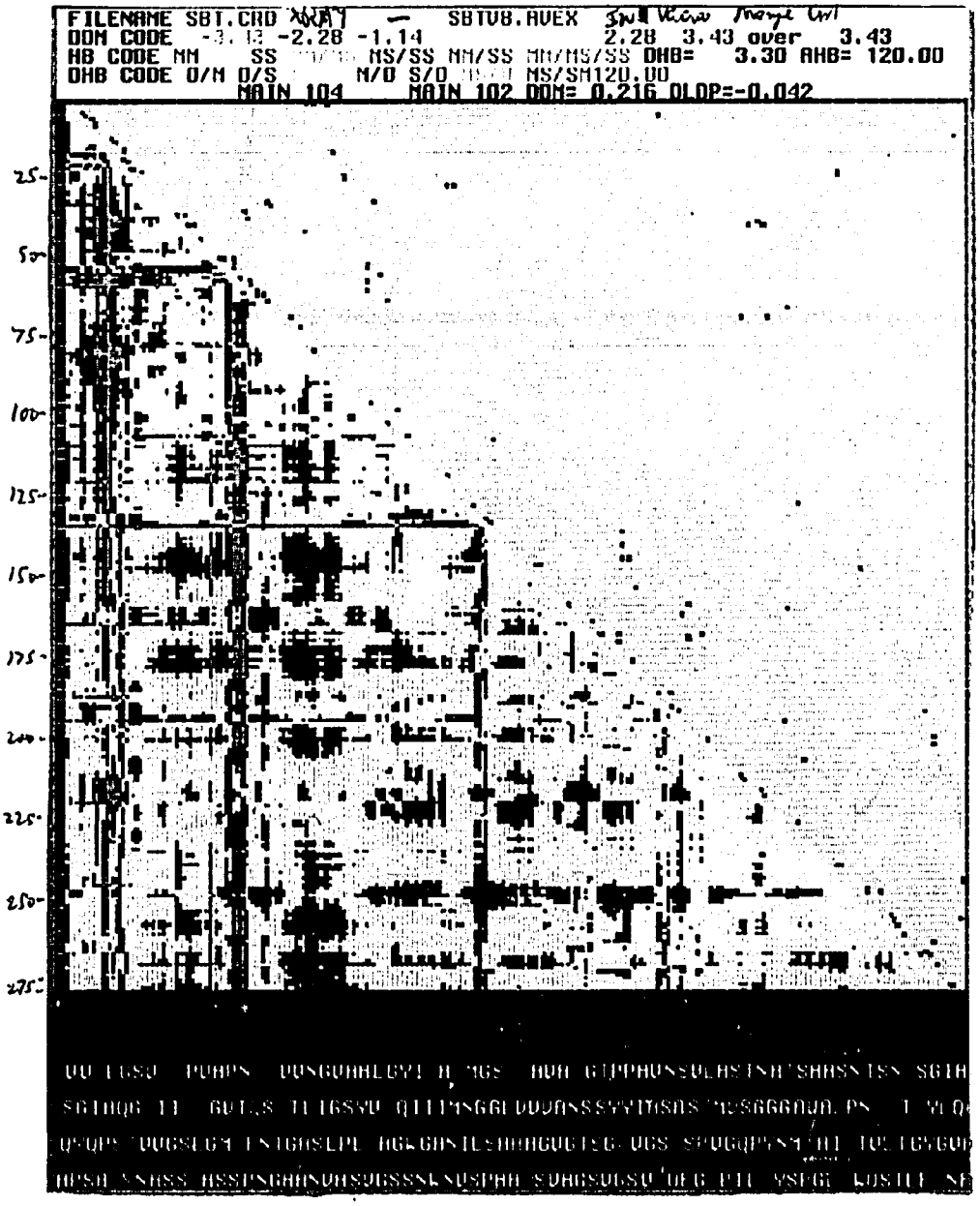


Figure 29a. DDM/DHBM plots of subtilisin xray structure vs. last 20ps average SBTV (*in vacuo*) structure. DDM is in bottom half of matrix, DHBM is in top half of matrix and primary sequence is at bottom of plot. Contours of DDM (in angstroms) are multiples of standard deviation (1.14Å); less than-3.43(orange), -3.43 to -2.28(dark red), -2.28to -1.14(red), -1.14 to 0.0(pink), 0.0 to 1.14(light blue), 1.14 to 2.28(blue), 2.28 to 3.43(dark blue) and greater than 3.43(black). HB cutoffs are 3.3Å between heavy atoms and 120 degrees for A-H-D angle. DHBM codes are O/M(red), O/S(blue), O/MS(yellow), M/O(dark red), S/O(dark blue), MS/O(orange) and MS/SM(purple) where O/X means gain and X/O means loss of hydrogen bond type X. Primary sequence residues are colored by polarity: non-polar(white), apolar(yellow), acidic(red) and basic(blue).

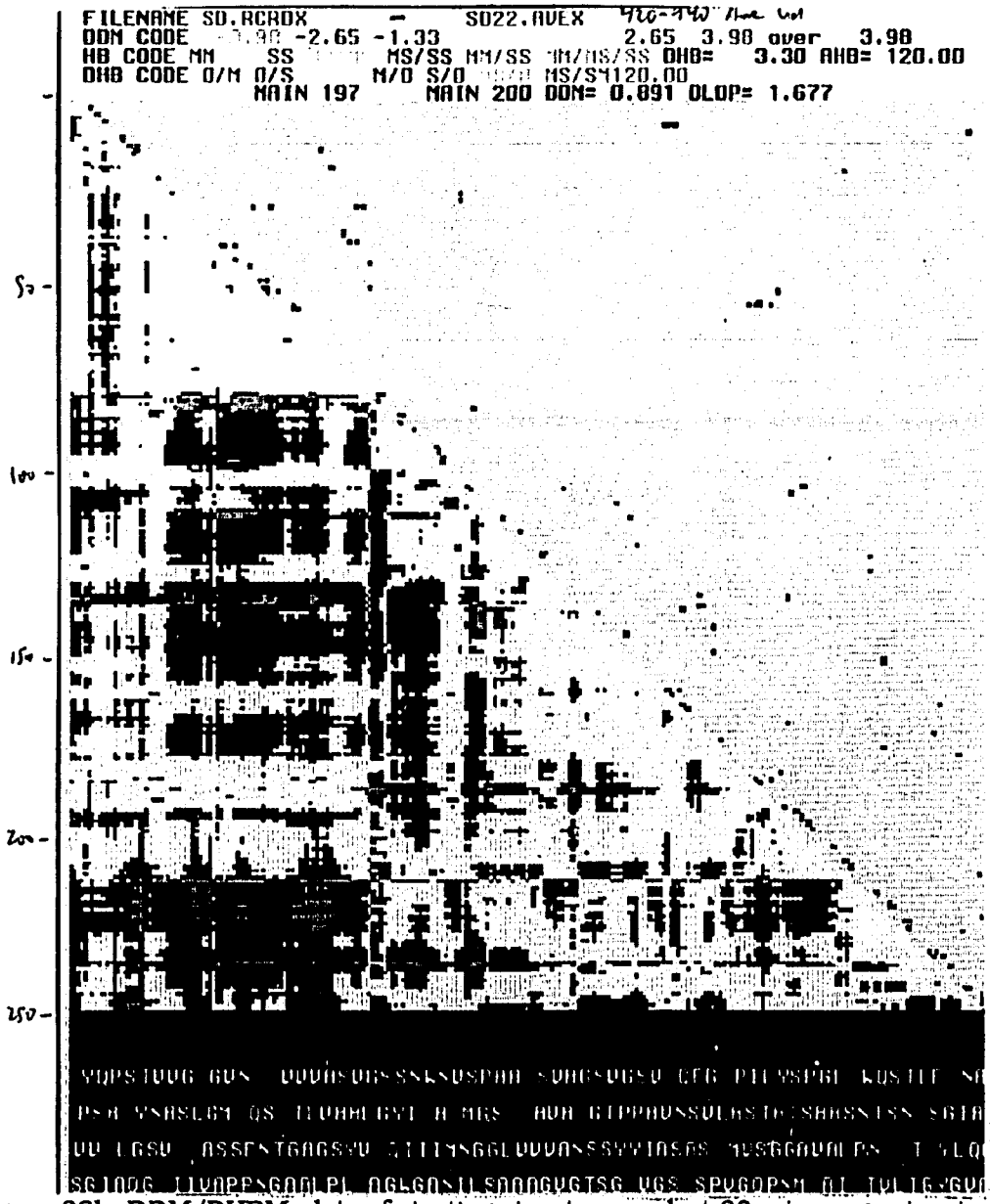


Figure 29b. DDM/DHBM plots of starting structure vs. last 20ps average structure from SBT30 (SBMD reduced dynamics simulation of subtilisin). There are only 258 residues in these structures; primary sequence numbers from native structure are shown in parentheses. DDM is in bottom half of matrix, DHBM is in top half of matrix and primary sequence is at bottom of plot. Contours of DDM (in angstroms) are multiples of standard deviation (1.33Å); less than-3.43(orange), -3.98 to -2.65(dark red), -2.65to -1.33(red), -1.33 to 0.0(pink), 0.0 to 1.33(light blue), 1.33 to 2.65(blue), 2.65 to 3.98(dark blue) and greater than 3.98(black). HB cutoffs are 3.3Å between heavy atoms and 120 degrees for A-H-D angle. DHBM codes are O/M(red), O/S(blue), O/MS(yellow), M/O(dark red), S/O(dark blue), MS/O(orange) and MS/SM(purple) where O/X means gain and X/O means loss of hydrogen bond type X. Primary sequence residues are colored by polarity: non-polar(white), apolar(yellow), acidic(red) and basic(blue).

structural rearrangement in the buffer is not important to the study of the solvated Ca^{2+} binding site in the reaction zone.

Table 5. Distances of oxygen ligands from Ca^{2+} in the *in vacuo* and solvated subtilisin simulations and the xray structure. Standard errors are in parentheses.

LIGAND	XRAY	SBT30	SBTV
GLY 169 O	2.66		2.49 (0.13)
TYR 171 O	2.97		2.41 (0.08)
PRO 172 O		2.40 (0.06)	2.80 (0.09)
VAL 174 O	2.66	2.62 (0.19)	2.37 (0.06)
GLU 195 O	3.02		
ASP 197 OD1	2.84	2.55 (0.09)	2.51 (0.07)
ASP 197 OD2		2.48 (0.06)	2.60 (0.10)
TIPS 42 OH2	2.72	2.37 (0.08)	2.39 (0.08)
TIPS 55 OH2		2.44 (0.08)	
TIPS 74 OH2	2.97	2.39 (0.06)	2.34 (0.06)
SOLV 703 OH2		2.39 (0.08)	

XRAY = Xray structure of subtilisin
SBT30 = Solvated subtilisin simulation
SBTV = *In vacuo* subtilisin simulation
TIPS = Crystal water
SOLV = Water added during solvation procedure

Table 5 shows the coordination distances of oxygen atoms to the Ca^{2+} in the simulations and the xray structure. The Ca^{2+} has a coordination number of 7 in the xray structure. The coordinating ligands for this low affinity binding site are larger than the normal 2.4Å in the xray structure and this reflects a low enthalpy of binding at this low affinity site (as the enthalpy of binding is a function of coordination distance). The optimized starting structures for both the *in vacuo* and solvated simulations have the same coordination number, 7, as the xray structure. However, the *in vacuo* optimized structure has significantly smaller coordination distances while the solvated optimized structure maintains coordination distances that are very similar to those in the xray Ca^{2+} binding site. Thus the presence of solvent maintains the

geometry of the coordination to the Ca^{2+} during optimization as well as the coordination number. The average structures resulting from both the *in vacuo* and solvated simulations have a coordination number of 8. The structures of these two Ca^{2+} binding sites are different both in coordinating ligands as well as in coordination geometry.

While there are only 7 ligands identified in the xray structure, it is possible that some water ligands are rapidly exchanging between the Ca^{2+} and the bulk solvent and are not observed in the xray structure. This is also suggested by the unusually large coordination distances observed in the xray structure. In support of this hypothesis, an exchange of coordination between two water molecules was observed in the solvated simulation near 520ps. A solvent molecule added in the solvation procedure (S1265) replaced one of the water ligands bound to the Ca^{2+} in the crystal structure (W55). This exchange also suggests that the coordination structure of the Ca^{2+} binding site is stable at this point in the simulation.

Figure 30 shows the rms alpha carbon positional difference, from the xray structure, in the solvated Ca^{2+} binding site. A conformational change occurred in this structure at ca. 230ps in the simulation. Figure 31 shows the time series of the coordinating ligands to the Ca^{2+} in the solvated simulation. Near to 230ps the aspartate carboxyl side-chain coordination to the Ca^{2+} becomes bidentate and replaces the backbone carbonyl ligand from glutamate 195. At ca. 300ps there is a change in coordination of loop 169-174 to the Ca^{2+}, observed in Figure 31, as a loss of coordination to the two backbone carbonyls of glycine 169 and tyrosine 171 and the addition of the carbonyl of proline 172 and one water molecule. The significant structural changes in the solvated Ca^{2+} binding site that occur at this point are the reorientation of the two protein loops and the replacement of one protein ligand (from loop 169-174) by a water ligand.

Fig. 30: RMS of CA atoms within 12.0 Angstroms
of Ca+2 in Binding Site B of SBMD Simulation of Subtilisin

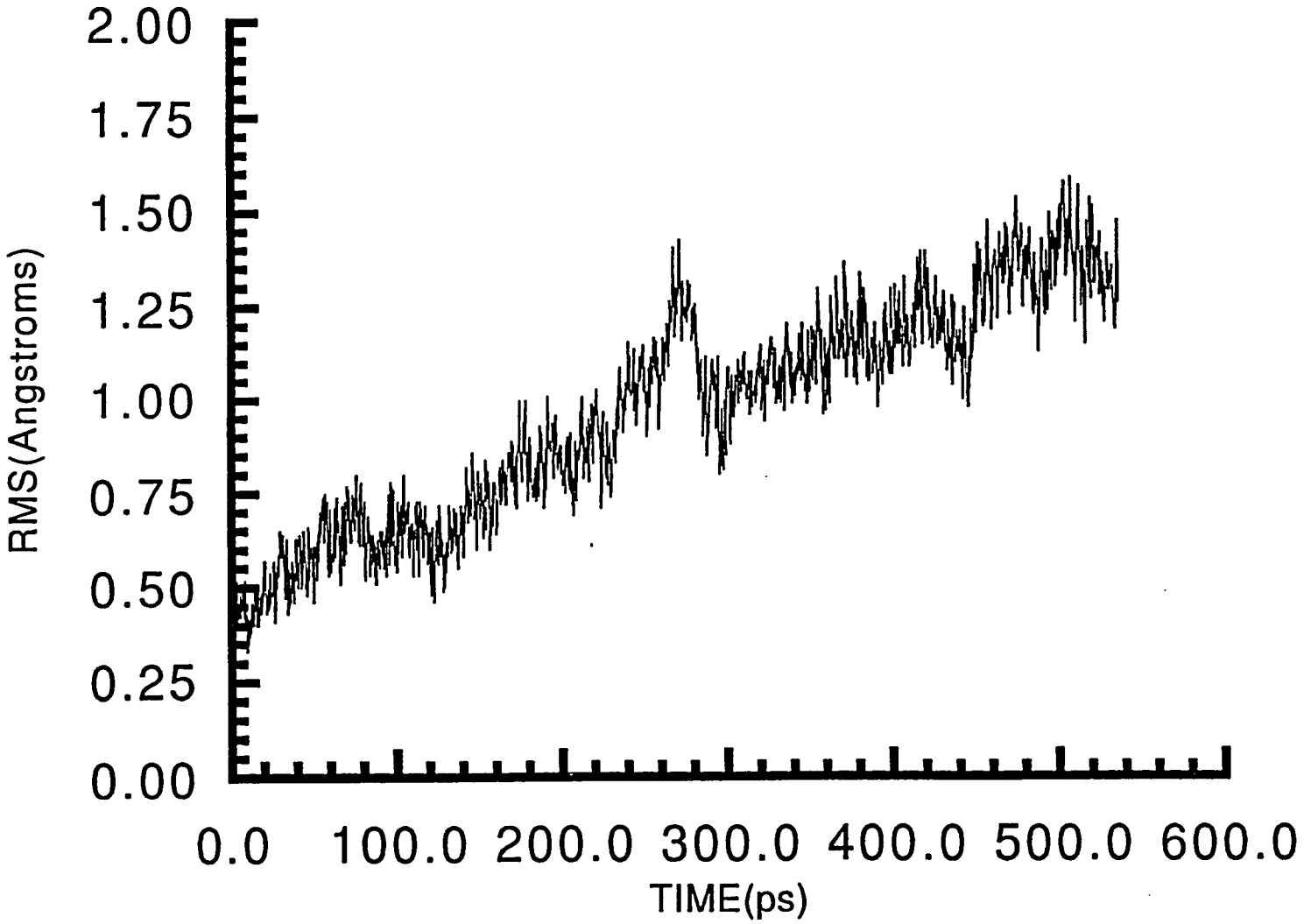


Fig. 31a: Time Series of Ca+2-Ligand Distances
from SBMD Subtilisin Simulation

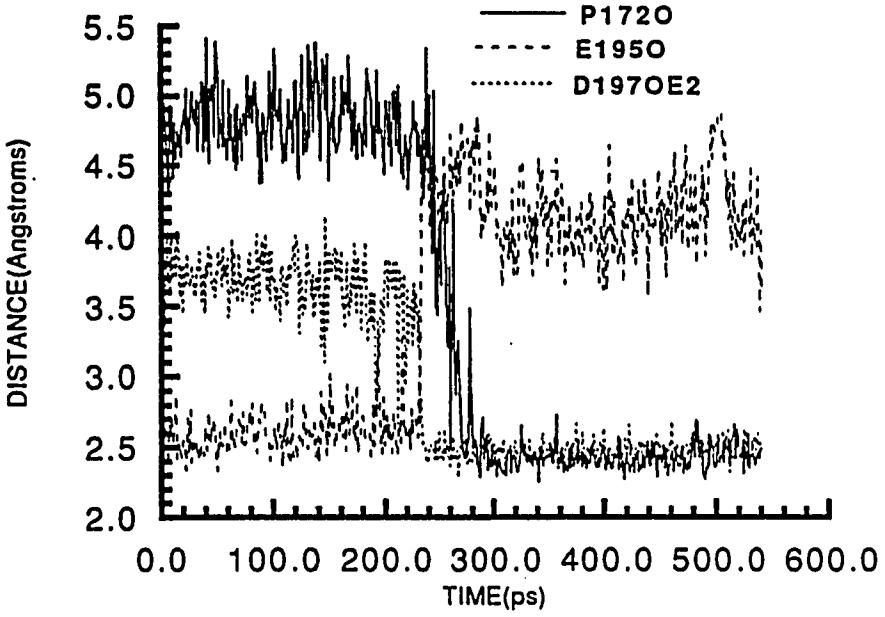


Fig. 31b: Time Series of Ca+2-Ligand Distances
from SBMD Subtilisin Simulation

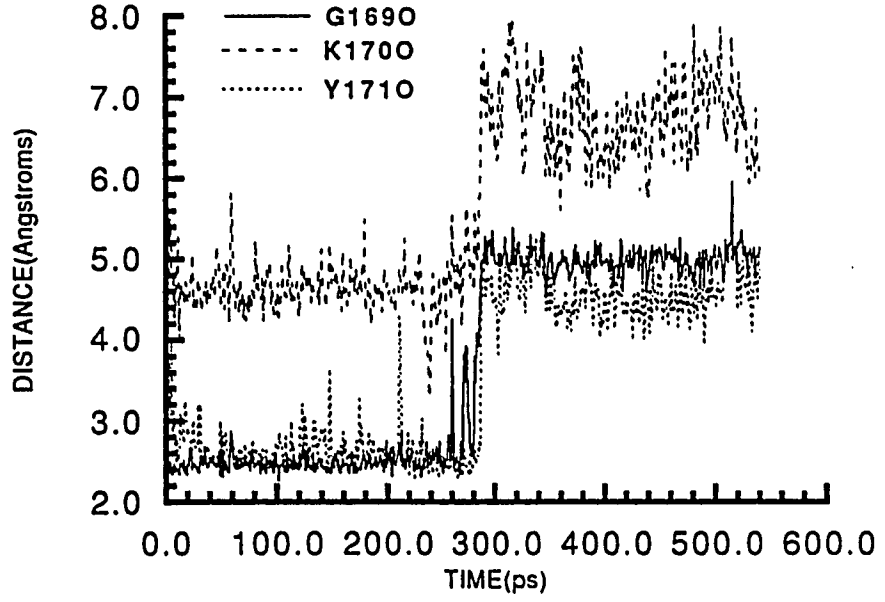


Fig. 31c: Time Series of Ca+2-Ligand Distances
from SBMD Subtilisin Simulation

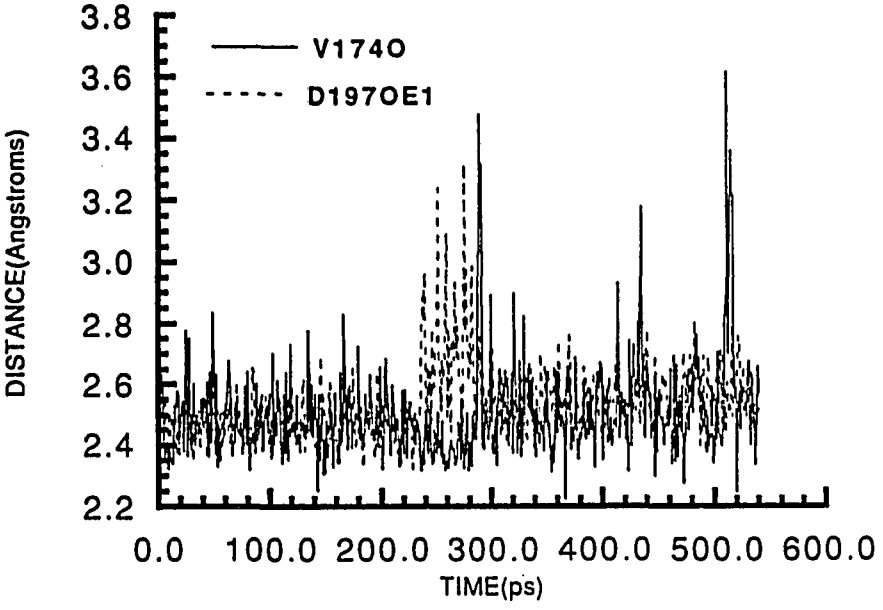
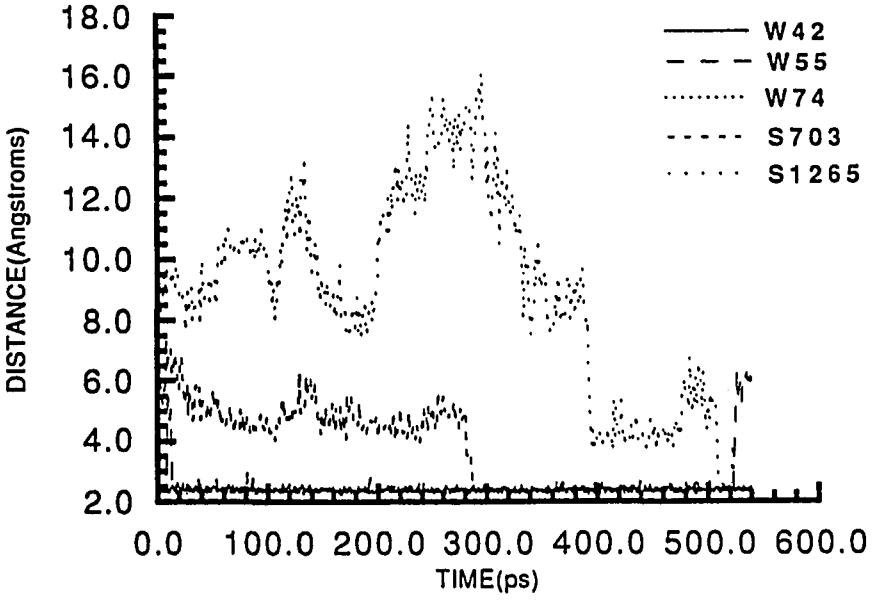


Fig. 31d: Time Series of Ca+2-Ligand Distances
from SBMD Subtilisin Simulation



Both simulations result in a multidentate coordination of the carboxyl group of the side-chain of aspartate 197. It is generally believed that the bidentate coordination results from the absence of polarization in the potential function. A significant amount of polarization is expected as a result of the interactions between dications and their ligands. The empirical potential functions we used do not include higher order electrostatic interactions (such as dipole-dipole interactions). In the absence of polarization the interactions of the carboxyl oxygens (e.g., on aspartate 197) with the Ca^{2+} is dominated by the coulombic (i.e. partial charge) term of the potential function, allowing these oxygen atoms to overcome their interatomic repulsion. This results in the bidentate coordination observed in these simulations.

Similarities to the changes observed in the coordination structure of the Ca^{2+} binding site in this solvated simulation (addition of several water ligands to the Ca^{2+} and reorientation of the protein loops) are observed in a simulation of the crystal structure of subtilisin [Heiner, 1992]. The bidentate coordination of the two oxygen atoms in the carboxyl of aspartate 197 was also observed in this simulation which used a different set of parameters and another computational program (the GROMOS program). The authors of that study also believe that the bidentate coordination forms as a result of overestimating electrostatic interactions (i.e. absence of polarization). While that simulation of the crystal structure was only over 60ps, the protein structures (4 in the unit cell of the crystal) were observed to have a drift in the rms positional difference of alpha carbons from the xray structure similar to the drift observed in the solvated simulation in this study (Figure 1b, Heiner, 1992); the crystal simulation resulted in a final rms difference of ca. 1.6Å after 60ps. The authors showed that the conformational change in the crystal simulation was due to both a deformation of the starting xray structure and an increase in the radius of gyration of the protein [Heiner, 1992]. The "blow up" was suggested to be the result of insufficient solvation of the

starting crystal structure. The major conformational changes in the solvated proteins of both studies are thus due to solvent effects.

In that simulation the starting structure was the xray structure of subtilisin with a K+ bound the to low affinity site B. When a simulation was performed with the K+ replaced with Ca^{2+} the coordination to the three ligands from loop 169-174 was lost and replaced by 5 waters for a total coordination number of 8. The simulation of the Ca^{2+} bound crystal structure of subtilisin started from the K+ bound xray structure of the binding site and thus would not be expected to yield the correct Ca^{2+} bound structure due to the short time scale of the simulation.

As the structure of the low affinity binding site is believed to be very flexible (Ca^{2+} affinity is 10^{-4}) it is not surprising that there were changes in the coordination of the Ca^{2+} during these solvated simulations. However in both the full crystal simulation and the reduced dynamics simulation of the solvated Ca^{2+} binding site the overall tertiary as well as secondary structure of the protein is maintained. The oscillatory departures of the protein structure from the xray structure observed in the *in vacuo* simulation are evidence that the molecular dynamics of the protein is qualitatively different in the absence of solvent. This suggests that the absence of solvent may destabilize or perturb the native conformation of the protein and that including solvent in the simulation of this binding site is important for accurately modelling the molecular dynamics of this system.

Solvent Effects on Protein Structure

The relationship between specific solvent networks identified in the xray structure of subtilisin, to the stability of the protein, was examined by relating changes in the structure of the solvent networks to changes in the structure of the solvated protein.

The crevice formed by the helices of residues 64-73 and 220-237 and the loop containing residues 202-217 is observed to contain a solvent network in the xray structure. This crevice is at the interface of the first domain and the other two domains, and in the solvated simulation it is much larger and more solvated than in the xray structure. As shown in the double HBM plot in Figure 32, the hydrogen bond interactions observed in the xray structure between the helix containing residues 64-73 and the loop containing residues 202-217 is absent due to the movement of these domains. Also, the structure of the protein around the helix containing residues 64-73 is where one of the largest conformational changes occurs in the *in vacuo* simulation. These results strongly suggest that solvent interactions contribute significantly to the stability of the local structure at this domain interface.

Another solvent network is found around the active site which is also a structure at a domain interface. The water bound to aspartate 32 in the active site of the xray structure is also observed in the solvated simulation. The structure of the active site in the solvated simulation is different however; whereas in the xray structure there is a hydrogen bond (at a distance of 2.4Å) between the side-chains of the active site residues serine 221 and histidine 64, in the average structure from the solvated simulation these side-chains are 7.2Å apart. In the *in vacuo* simulation these side-chains are 4.1Å apart. This structure is formed at the interface between the two domains whose relationship is altered by the movement of the reaction zone toward the solvent while the N-terminal domain in the buffer region is constrained.

Solvent exerts major effects on the protein. The major conformational changes in tertiary and secondary structure observed in the reduced dynamics solvated simulation result from the boundary conditions restricting the interaction of the N-terminal domain with the solvent. The absence of solvent, shown by the results of the *in vacuo* simulation, significantly perturb the dynamical properties of the molecule and also restrict the conformational flexibility of the Ca^{2+} binding site. Thus solvent

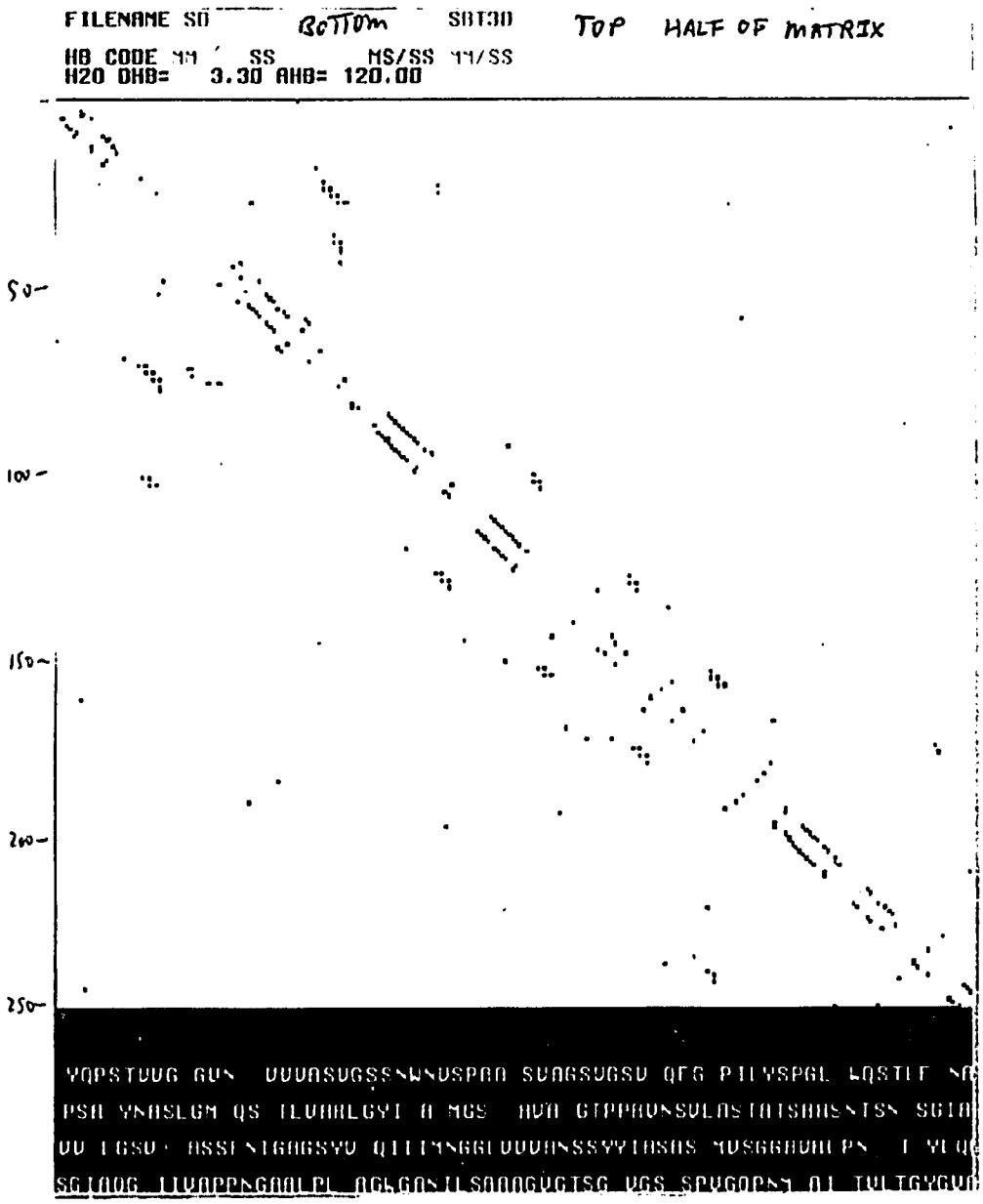


Figure 32. HBM plots of starting structure vs. last 20ps average structure from SBT30 (SBMD reduced dynamics simulation of subtilisin). HBM plot of SBT30 structure is in the top half of the matrix and HBM of the starting structure is in the bottom half of the matrix. There are only 258 residues in these structures; primary sequence numbers from native structure are shown in parentheses. HB cutoffs are 3.3Å between heavy atoms and 120 degrees for A-H-D angle. HBM colors are MM(red), MS(yellow), SS(blue), MM/MS(orange), MS/SS(green), MM/SS(purple) and MM/MS/SS(brown). Primary sequence residues are colored by polarity: non-polar(white), apolar(yellow), acidic(red) and basic(blue).

effects were included in the following study on the role of serine 163 in the Ca^{2+} binding site.

The major artifact observed in the reduced dynamics simulation of the Ca^{2+} binding site of subtilisin, the conformational change in the interaction of the buffer zone and the reaction zone, show that caution must be used in drawing conclusions from these results to learn about the structure and properties of the enzyme function of the protein. Specifically, any conclusions about the active site and substrate binding site of the protein must be considered questionable in light of the artifact observed in the simulation. The conclusions we wish to draw, however, concern the Ca^{2+} binding site which is at the center of the reaction zone of this simulation. If the subject of this study was the enzyme function of the protein it would have been appropriate to chose the active site as the center of the reaction zone of the simulation and, perhaps, these structural changes (at the active site) would not have occurred. These results emphasize the limitation of reduced dynamics simulations as being applicable to questions that can be localized to structures that can be centered within a reaction zone.

D. Mutation of Serine 163 and Effects on Ca^{2+} Binding Site B in Subtilisin

The role of serine 163 in the mechanism of Ca^{2+} binding to the low affinity site in subtilisin was explored by comparing the structure and dynamical properties of the native solvated Ca^{2+} binding site with the properties of mutants constructed computationally. This study used the equilibrated native structure of the solvated Ca^{2+} binding site obtained from the reduced dynamics (SBMD) simulation of the previous comparison of subtilisin in solvent and *in vacuo*.

Serine 163 was mutated to an alanine using two computational methods described below: a free energy perturbation and a sudden deletion procedure. The Ca^{2+} free structures of the native and mutant proteins were constructed using the sudden deletion procedure only as it was not feasible to construct the Ca^{2+} free structures using

the perturbation method (see below). Ca^{2+} bound simulations of mutants constructed with the sudden deletion procedure were used to check the accuracy of this method against the more reliable perturbation procedure. The hypothesis that the serine has a role in determining the conformation of the Ca^{2+} binding site, through its effect on the hydrogen bonding interactions between residues in the protein loops containing the coordinating ligands, was studied by identifying changes in the hydrogen bonding interactions resulting from the mutation Ser163Ala.

The detailed analysis described below shows that the changes in the secondary and tertiary structure resulting from the mutations were not as pronounced as the changes in the dynamics properties of the protein. The mutation resulted in conformational changes which were observed in the LDP and DM plots to be different from the changes resulting from the absence of Ca^{2+}. Differences in hydrogen bonding interactions between the mutant and native structures of the Ca^{2+} binding sites were also observed, and led to the hypothesis that these differences in hydrogen bonding interactions produce the changes in structure and dynamics observed in the mutant and apo proteins.

The structural rearrangement following the mutation of serine 163 described below, resulted in a hydrogen bond between the side chain of threonine 164 and the main chain carbonyl of serine 162. This was accompanied by a conformational change in the interaction of serine 191 with these residues. In the computed apo native structures this hydrogen bond interaction, between residues 162 and 164 is not observed. The result of this hydrogen bond interaction is a difference in the response that residues 186 to 192 exhibit to the absence of Ca^{2+}, and in a structural change in their interactions with residues 154 to 158. This difference in the interactions between the loop containing serine 163 and the loop containing residues 186-192 is identified here as the mechanism by which the mutation affects the dynamics properties of the Ca^{2+} binding site.

In the apo native protein, compared to the Ca^{2+} bound structure, a difference was observed in the tertiary structural interactions between the loop containing the mutation (residues 155 to 165), the loop containing residues 180 to 190, and the loops of the Ca^{2+} binding site. Moreover, significant differences emerged in the molecular dynamics properties expressed by the apo and Ca^{2+} bound structures. This is in contrast to the results observed in the mutant structures. The structure and dynamics of the Ca^{2+} binding site in the apo and Ca^{2+} bound mutants did not differ significantly.

Structural changes in the substrate binding site were also observed in the apo structures of both the native and mutant proteins, compared to their Ca^{2+} bound structures. This suggests that the modulation of protein activity by Ca^{2+} binding is through effects on the substrate binding site.

Table 6. List of simulations performed in the Serine 163 mutation study of subtilisin.

SBT30	=	440ps structure from SBMD simulation of subtilisin Ca^{2+} binding site B (starting structure).
SBT30F	=	100ps simulation starting from SBT30 structure with Ca^{2+} deleted.
S163A	=	Ser163Ala perturbation simulation starting from SBT30 structure (157.5ps).
SA	=	100ps simulation starting from final S163A structure.
SAF	=	SA simulation with Ca^{2+} deleted.
SP	=	100ps sudden deletion simulation of Ser163Ala mutant starting from SBT30 structure.
SPF	=	SP simulation with Ca^{2+} deleted.

Construction and Equilibration of S163A Mutant and Apo Protein Structures

Free energy Perturbation Method, Sudden Deletion Method

The simulations performed in this study of the role of serine 163 are listed in Table 6. The mutation of serine 163 to alanine in a free energy perturbation procedure started from the structure at 440ps obtained from the solvated simulation of the Ca^{2+} binding site B in subtilisin, termed SA. The windowing free energy perturbation procedure as implemented in CHARM [Brooks, 1983b] (described in the Methods section) was used. The window intervals were 2.5ps of equilibration followed by 5ps of production dynamics. The step size (perturbation parameter) between windows was calculated from a predetermined limit of app. 1.0 +/-0.5 kcals in the free energy change between intervals. Using a free energy change limit of app. 1.0kcal per step resulted in 21 windows being required for the entire perturbation. The structure resulting from this procedure was then equilibrated for 100ps.

The sudden deletion procedure for constructing the alanine mutant was a simple computational deletion of the hydroxyl side chain of the serine. Since extended atom representations of non-polar hydrogens were used in these simulations (see Background section) the serine beta carbon extended atom (which represents a carbon and hydrogen atom) and the side chain atoms were replaced by an alanine beta carbon extended atom (which represents a carbon and two hydrogens). The sudden deletion mutant was constructed from the same starting structure used in the SA simulation and was equilibrated for 100ps. This simulation was called SP.

Ca^{2+} free structures were constructed using the sudden deletion method. The free energy perturbation method was not considered for constructing the Ca^{2+} free structures as the free energy change in the perturbation simulation (to the Ca^{2+} free structure) is on the order of several hundred kcals (and would require on the order of thousands of CPU hours). A Ca^{2+} free structure of the SP alanine mutant was

constructed, called SPF, and a Ca^{2+} free structure of the SA alanine mutant was constructed and called SAF. A Ca^{2+} free native protein structure was also constructed from the same starting structure used to make the alanine mutants (the 440ps structure from the previous solvated simulation). This simulation was called SBT30F. Each of these structures were equilibrated for 100ps.

The temperature and potential energy plots obtained from the simulations described above (Figures 33 and 34) show that the systems appear to have reached thermal equilibration.

Comparison of Native and Mutant Subtilisin Structures

A. RMS Differences

The average structures obtained from the last 20ps interval of each simulation are used to present the structural changes among the mutants and apo proteins. Likewise, the dynamics properties presented are from the last 20ps interval of each simulation. The results of the Ca^{2+} bound mutant simulations using the two different methods were similar.

The alpha carbon positional differences among the average structures from the mutant and apo subtilisin simulations and the native structure are shown in Table 7. The rms alpha carbon positional difference, from the native structure, to the apo native structure (0.96Å) is similar to the differences resulting from the mutation relative to the starting structure (1.21Å to SA and 0.91Å to SP average structures); the greatest difference from the native structure is to the apo perturbation mutant (1.45Å to SAF). The difference between the apo native structure and both the apo and Ca^{2+} bound mutants is much larger. The rms difference from the apo native structure (SBT30F) is 1.68Å for SAF and 1.50Å for SPF structures. Thus the structural difference between the apo native structure and the mutant structures (both apo and Ca^{2+} bound) is much greater than the difference between any of these structures from the Ca^{2+} bound native

Fig 33a: Temperature Plot from Simulation of Apo Native Subtilisin

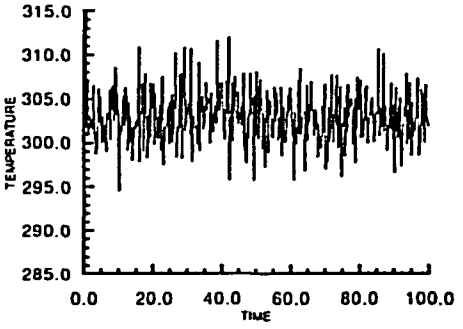


Fig. 33b: Temperature Plot from Perturbation Mutation Simulation of Subtilisin

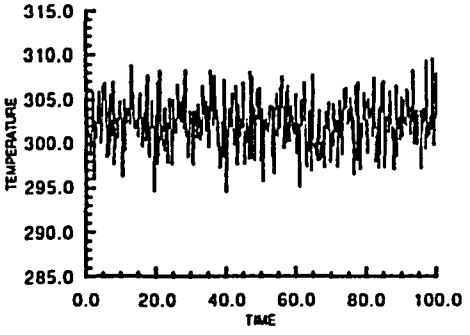


Fig 33c. Temperature Plot from Apo Perturbation Mutation Simulation of Subtilisin

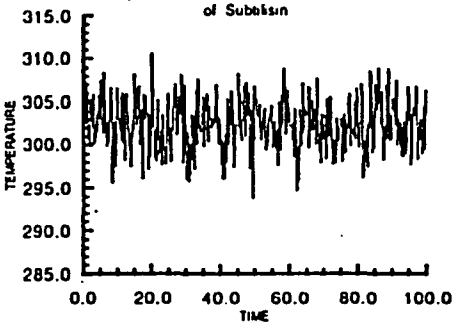


Fig. 33d: Temperature Plot from Sudden Deletion Mutation Simulation of *Subtilisin*

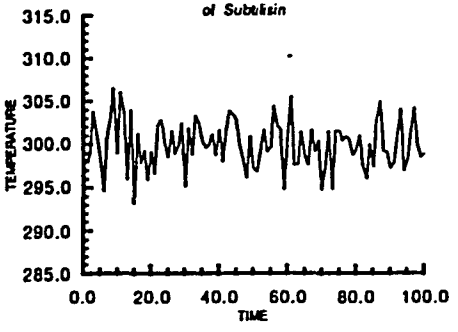


Fig 33e: Temperature Plot from Apo Sudden Deletion Mutation Simulation of Subtilisin

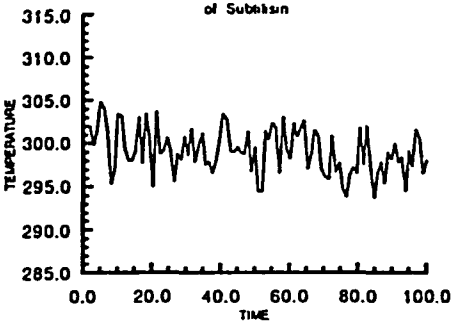


Fig 34a Potential Energy Plot from Simulation of Apo Native Subtilisin

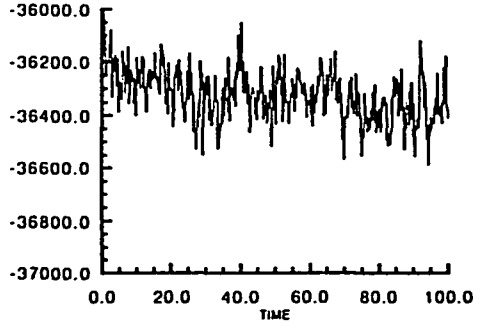


Fig. 34b: Potential Energy Plot from Perturbation Mutation Simulation of Subtilisin

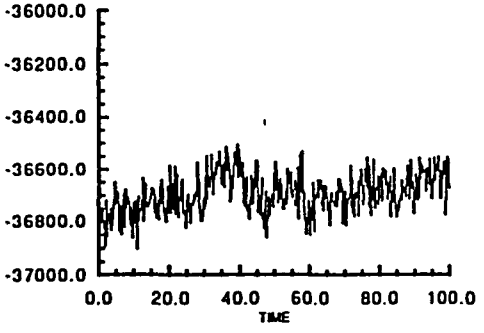


Fig 34c Potential Energy Plot from Apo Perturbation Mutation Simulation of Subtilisin

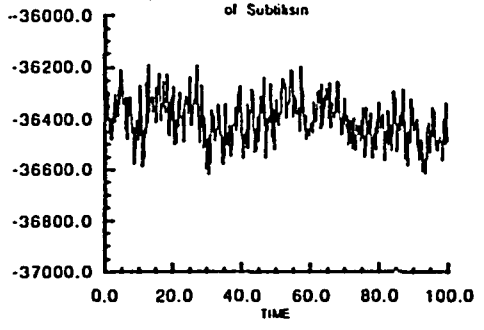


Fig. 34d: Potential Energy Plot from Sudden Deletion Mutation Simulation of Subtilisin

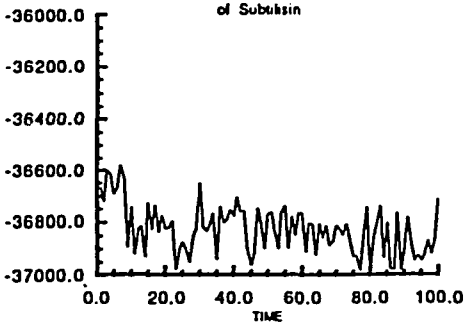
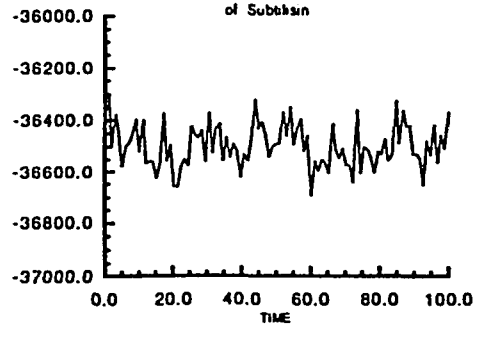


Fig 34e Potential Energy Plot from Apo Sudden Deletion Mutation Simulation of Subtilisin



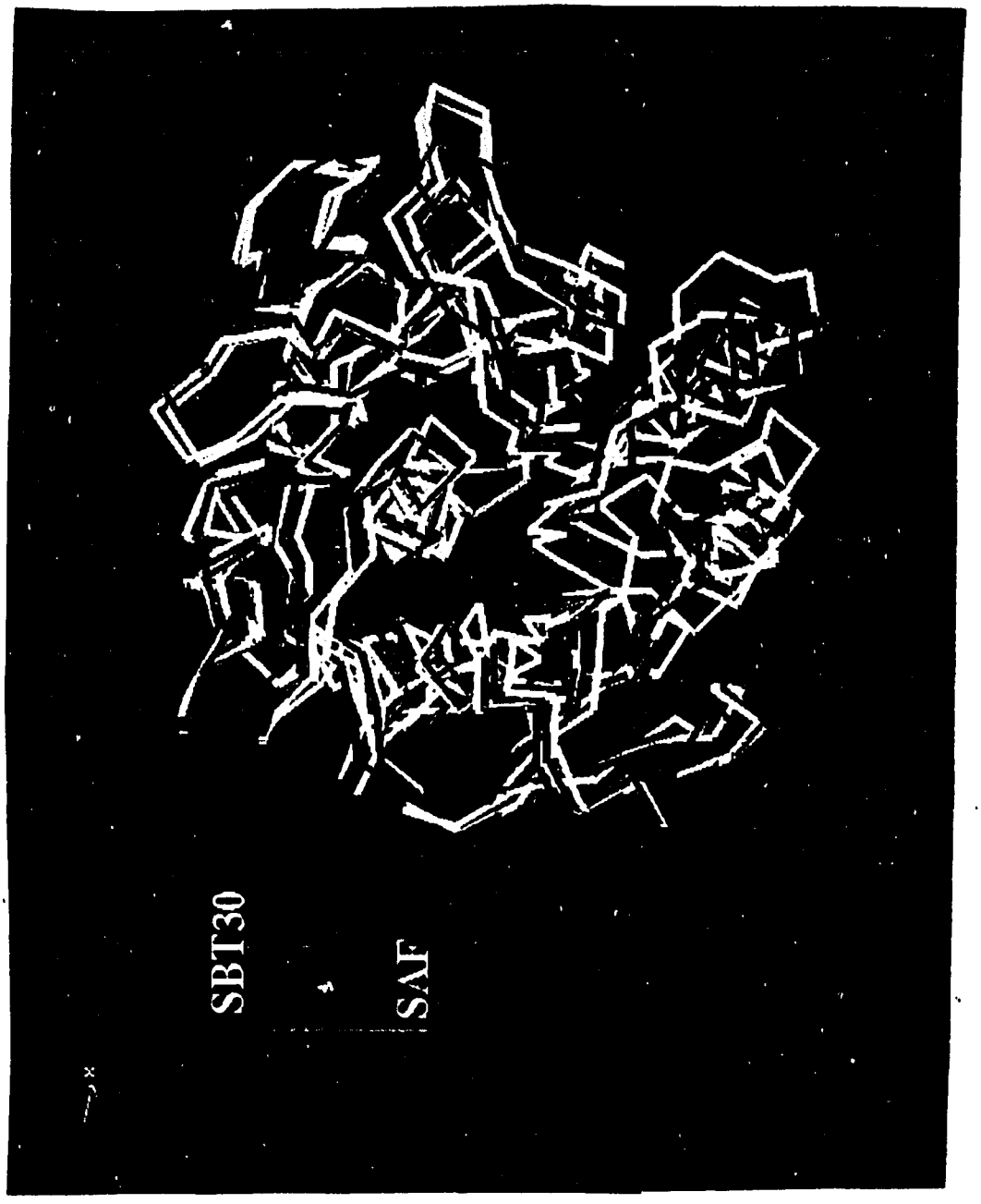


Figure 35. Alpha carbon plots of Ca^{2+} bound and apo native and perturbation mutant subtilisin structures. Color code is; Ca^{2+} bound native (SBT30 average structure, white), apo native (SBT30F average structure, red), Ca^{2+} bound mutant (SA average structure, blue), and apo mutant (SAF average structure, yellow).

Table 7. Alpha carbon rms positional differences among the last 20ps average structures from the mutant and apo subtilisin simulations and the xray structure.

Crds	SBT30	SBT30F	SA	SAF	SP	SPF
SBT30	0.00					
SBT30F	0.96	0.00				
SA	1.21	1.32	0.00			
SAF	1.45	1.68	1.14	0.00		
SP	0.91	1.48	1.21	1.42	0.00	
SPF	0.82	1.50	1.00	1.26	1.04	0.00

SBT30 = 440ps structure from SBMD simulation
SBT30F = SBT30 simulation with Ca^{2+} deleted
SA = S163A perturbation simulation
SAF = SA simulation with Ca^{2+} deleted
SP = S163A sudden deletion simulation
SPF = SP simulation with Ca^{2+} deleted

protein. This suggests a difference between the structural changes resulting from the absence of Ca^{2+} in the native structure and the structural changes resulting from the mutation.

Figure 35 shows the alpha carbon plot of the conformation of the loop containing the mutated residue (residues 155-165) in both the Ca^{2+} bound and apo native and perturbation structures. The apo native structure is the only one in which the tertiary conformation of the loop differs from the native structure. As indicated by the shift in position of the loop in the apo native structure there are changes in the structural interactions with adjacent loops (specifically residues 186-192) as well as changes in side chain conformations within this loop that are not shown by the alpha carbon plot. These differences are described below.

B. Secondary Structure

The LDP plots of the mutant and apo structures and the native structure are shown in Figure 36. The difference observed among these plots at residues 75 to 85 is in a segment of the protein structure that is adjacent to the buffer zone of the SB procedure

(residues 76 to 80 are in the reservior and were deleted from the structure); the discontinuity at residue 211 is due to the removal of residues 211-214 from the reservior.

The DLDP plot shown in Figure 37 compares the secondary structures of the Ca^{2+} bound and apo structures resulting from the perturbation and sudden deletion procedures. The two mutation procedures resulted in similar changes in both the Ca^{2+} bound and apo simulations; apart from the local structure around residues 70 to 80 and residue 240 the protein secondary structures resulting from the two methods are similar.

The DLDP plot of the average structure from the Ca^{2+} bound perturbation simulation, of the native structure, is shown in Figure 38. A conformational change is observed at residues 80 to 90 and in the segment around residue 240. The DLDP plot shown in Figure 39 compares the differences between Ca^{2+} bound and apo structures of the native and perturbation simulations. This shows that the differences observed (Ca^{2+} bound vs. apo structure) in the secondary structure of the native protein are greater than those observed in the mutant. Apart from the differences observed in both the native and mutant structures at residues 75 to 85 there are differences between the native Ca^{2+} bound and apo structures in the loop containing the mutation, residues 155 to 165, and around residues 210 and 240.

C. Tertiary Structure

The DM plots of the native structure and apo native structure are shown in Figure 40 and the Ca^{2+} bound and apo mutant structures from the perturbation simulations are shown in Figure 41. The DDM plots in Figure 42 show the changes in tertiary structure, compared to the native structure, resulting from the perturbation simulations and the apo native simulation.

The DDM plot of the Ca^{2+} bound vs. apo native structure shows structural differences between the first N-terminal domain and the two carboxy terminal

Fig. 36a: LDP of Starting Structure for SBMD Solvated Simulations of Subtilisin

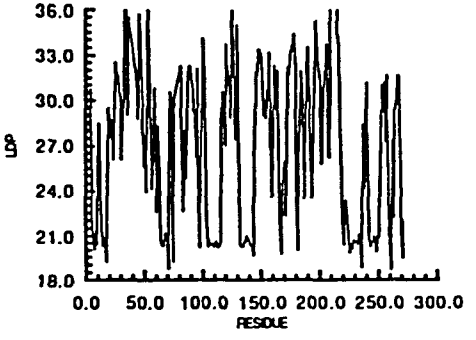


Fig. 36a LDP from Simulation of Apo Native Subtilisin

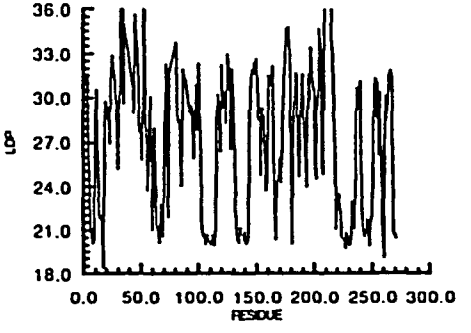


Fig. 36b: LDP from Perturbation Mutation Simulation of Subtilisin

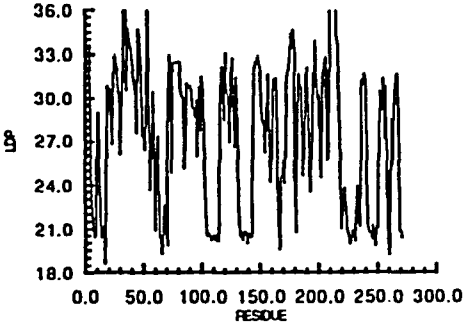


Fig 36c: LDP from Apo Perturbation Mutation Simulation of Subtilisin

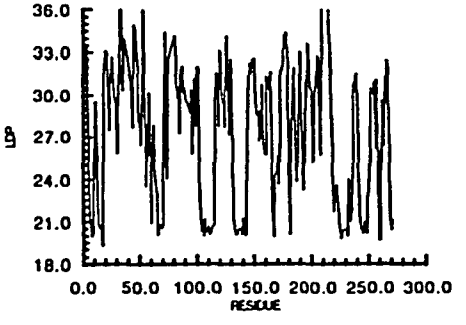


Fig. 36d: LDP from Sudden Deletion Mutation Simulation of Subtilisin

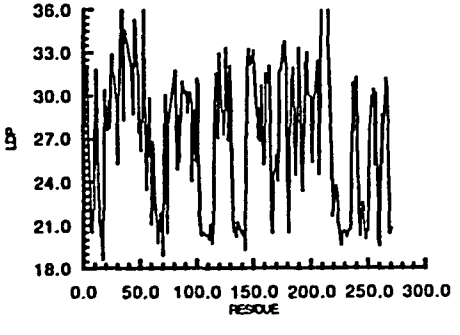


Fig 36e. LDP from Apo Sudden Deletion Mutation Simulation of Subtilisin

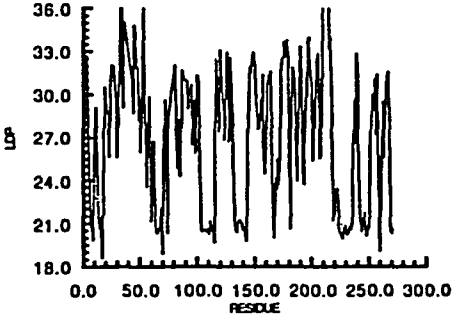


Fig. 37: DLDP of Perturbation vs. Sudden Deletion Mutants from Subtilisin Simulations

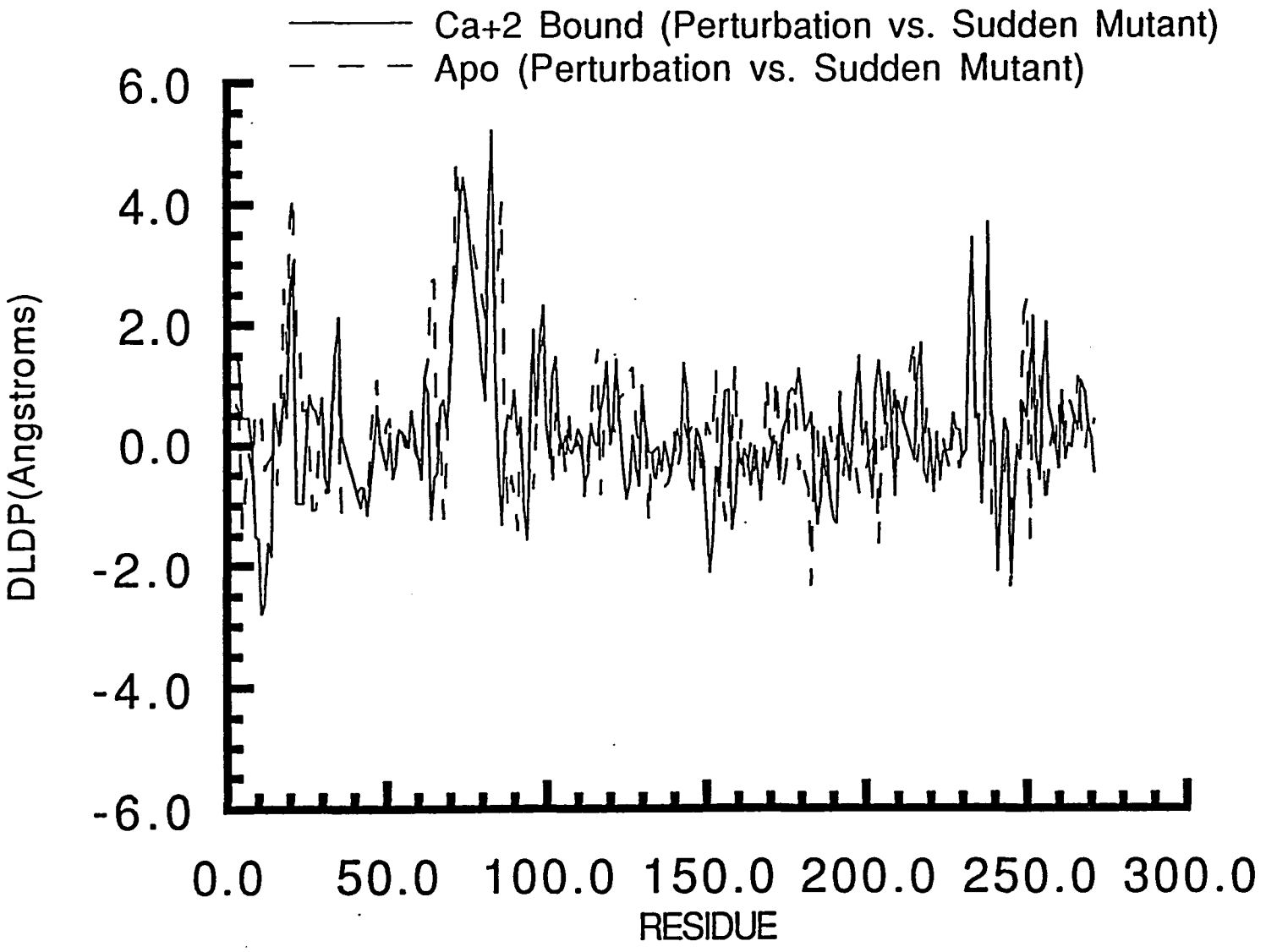


Fig. 38: DLDP of Ca+2 Bound
Perturbation Mutation Structure vs. Starting Structure

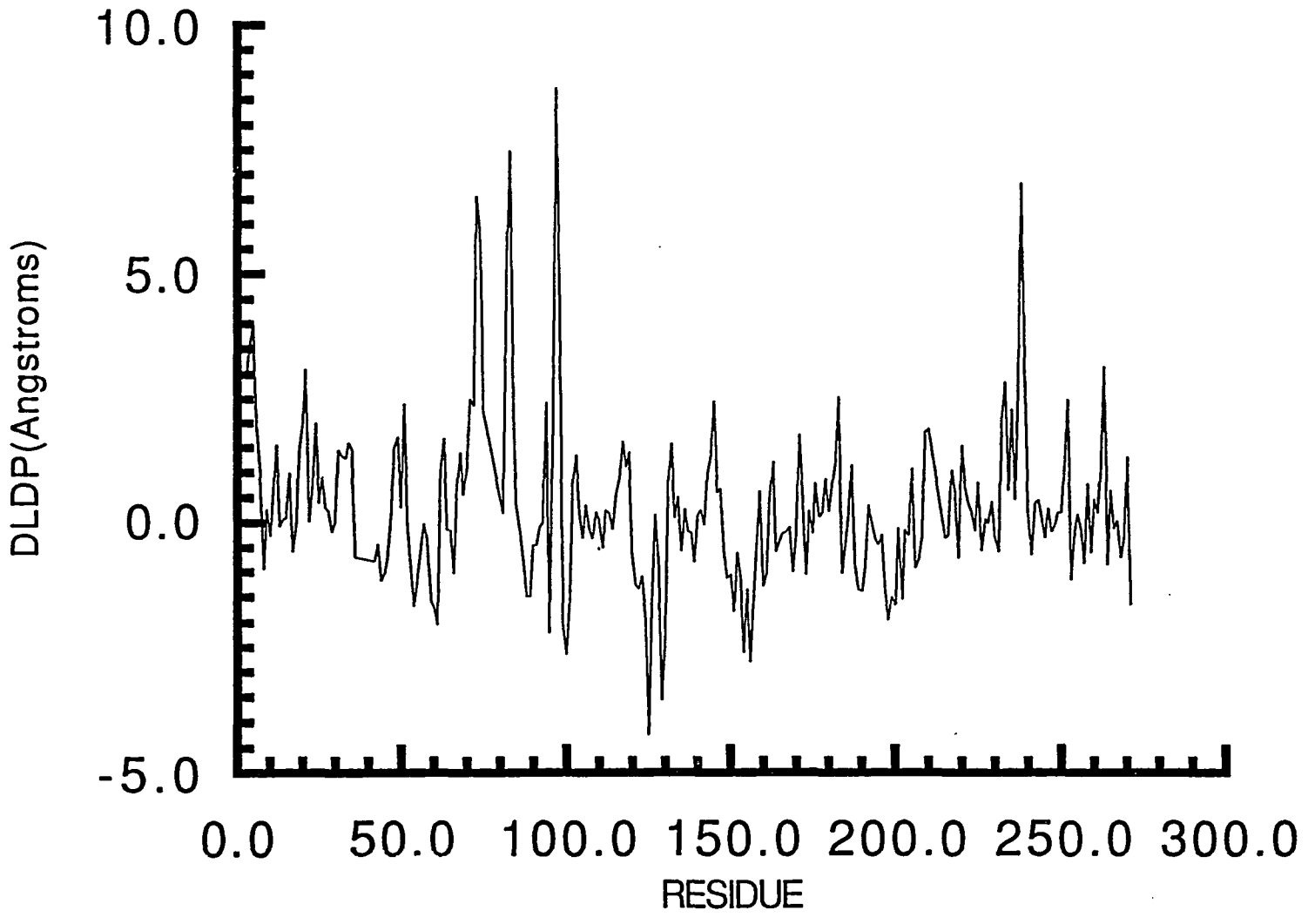
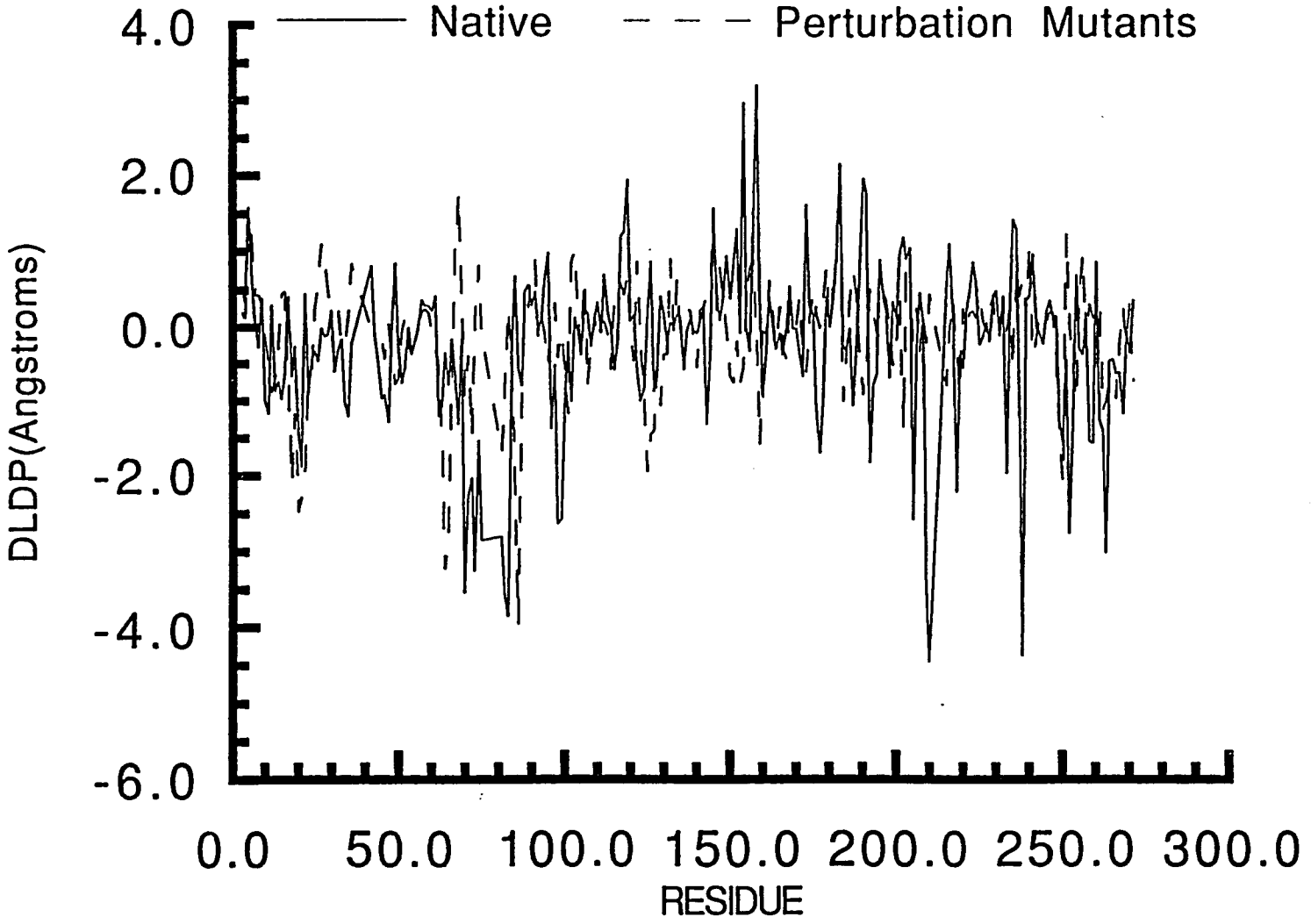


Fig. 39: DLDP of Ca+2 Bound vs. Apo
Average Structures from Native and
Perturbation Mutation Subtilisin Simulations



domains; the change, centered around residue 100, is similar to that observed in the solvated simulation of native subtilisin in the previous study (see Results section C). There is a difference in the tertiary conformation of the loop containing the mutation, residues 155 to 165, identified in the DDM plot as blue areas between this loop (155-165) and residues 180 to 190, and residues 200 to 215. In the apo structure the loop containing the mutation (155 to 165) is closer to these other structures compared to its position in the Ca^{2+} bound native protein. A tertiary conformational change around residue 240 is identified by the blue horizontal line extending virtually the entire length of the bottom of the plot.

The DDM plots of the native vs. the mutant structures are significantly different. Again there is a small change around residue 100 but the striking difference is the conformation of the carboxyl terminal domains. There are numerous changes in the conformation of the mutant structure throughout the portion including residues 160 to 275, compared to the same region in the native structure. These changes are observed in the DDM plots of both the Ca^{2+} bound and apo mutant vs. native structures and are thus the result of the mutation and not the absence of Ca^{2+}. The mutation resulted in changes in the conformation of the carboxy terminal domain and in the interactions between the loop containing residues 155 to 165, and the loops containing residues 180 to 190, and 190 to 200. These structural changes are associated with changes in the dynamics properties of the mutant and native proteins observed in the CDM plots (discussed below).

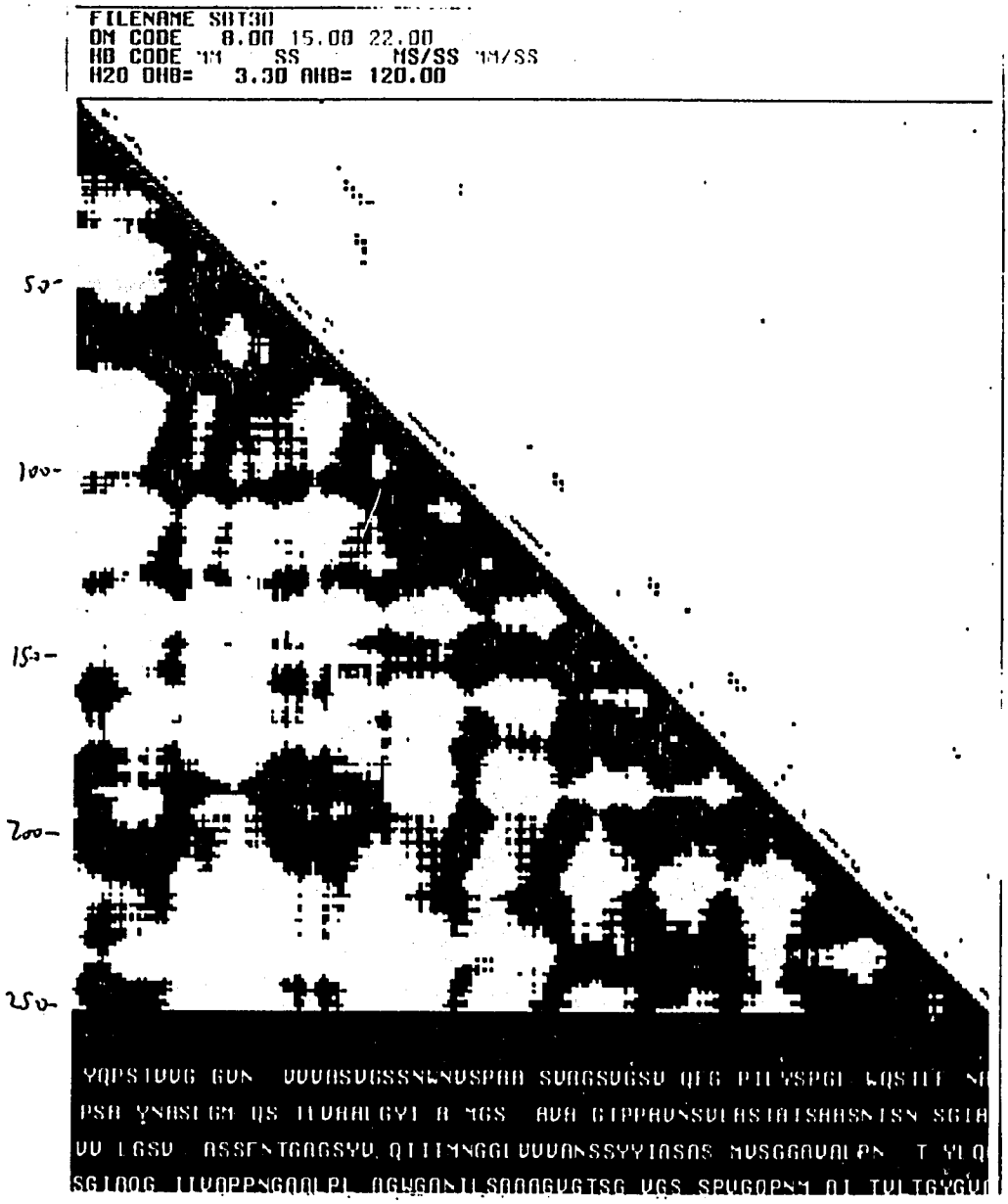


Figure 40a. DM/HBM plots of last 20ps average structure from SBT30 (Ca^{2+} bound native) SBMD reduced dynamics simulation of subtilisin. There are only 258 residues in this structure; primary sequence numbers from native structure are shown in parentheses. DM is in bottom half of matrix, HBM is in top half of matrix and primary sequence is at bottom of plot. Contours of DM are 0-8Å(green), 8-15Å(light red), 15-22Å(dark red) and 22-30Å(orange). HB cutoffs are 3.3Å between heavy atoms and 120 degrees for A-H-D angle. HBM colors are MM(red), MS(yellow), SS(blue), MM/MS(orange), MS/SS(green), MM/SS(purple) and MM/MS/SS(brown). Primary sequence residues are colored by polarity: non-polar(white), apolar(yellow), acidic(red) and basic(blue).

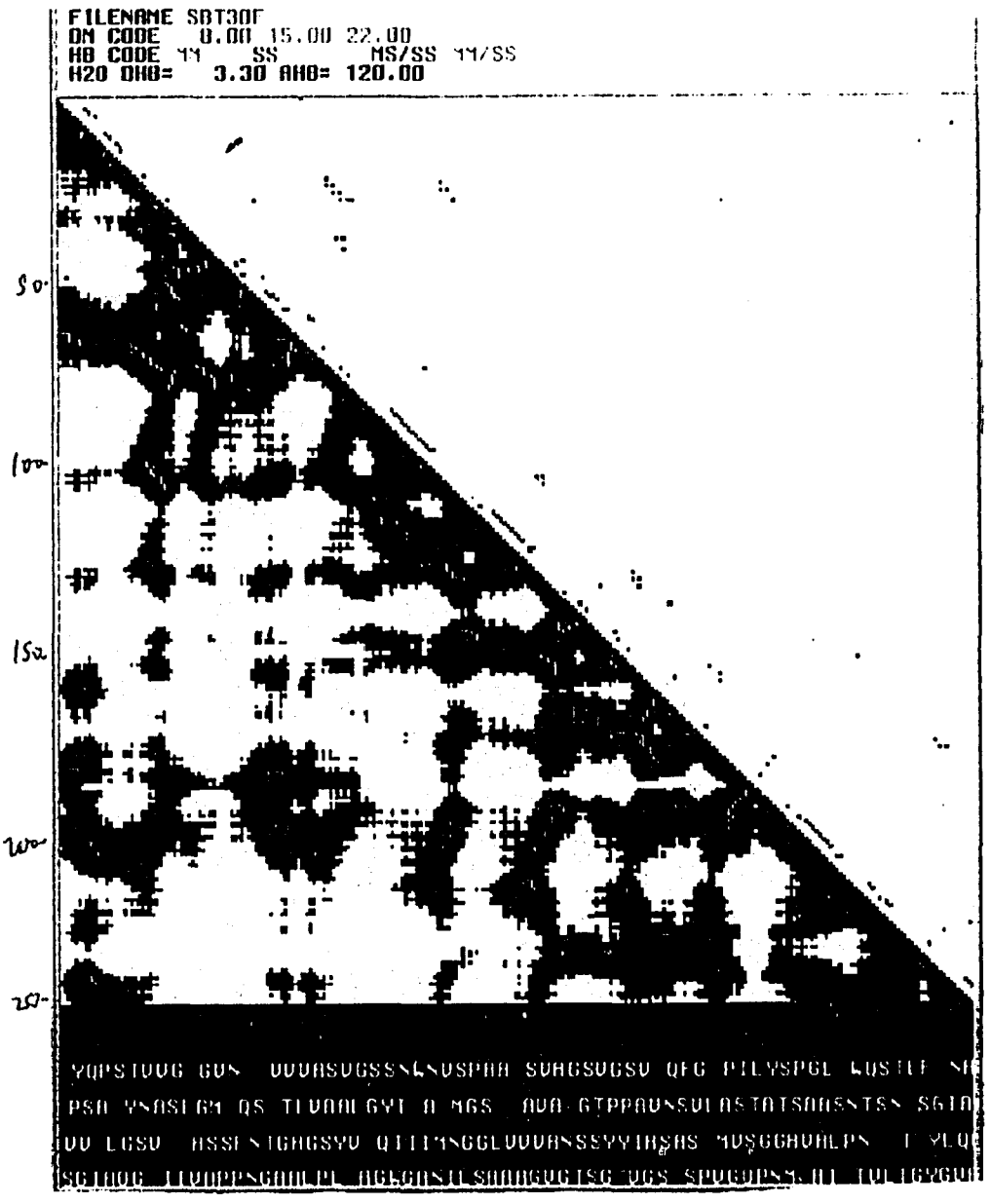


Figure 40b. DM/HBM plots of last 20ps average structure from SBT30F (apo native) SBMD reduced dynamics simulation of subtilisin. There are only 258 residues in this structure; primary sequence numbers from native structure are shown in parentheses. DM is in bottom half of matrix, HBM is in top half of matrix and primary sequence is at bottom of plot. Contours of DM are 0-8Å(green), 8-15Å(light red), 15-22Å(dark red) and 22-30Å(orange). HB cutoffs are 3.3Å between heavy atoms and 120 degrees for A-H-D angle. HBM colors are MM(red), MS(yellow), SS(blue), MM/MS(orange), MS/SS(green), MM/SS(purple) and MM/MS/SS(brown). Primary sequence residues are colored by polarity: non-polar(white), apolar(yellow), acidic(red) and basic(blue).

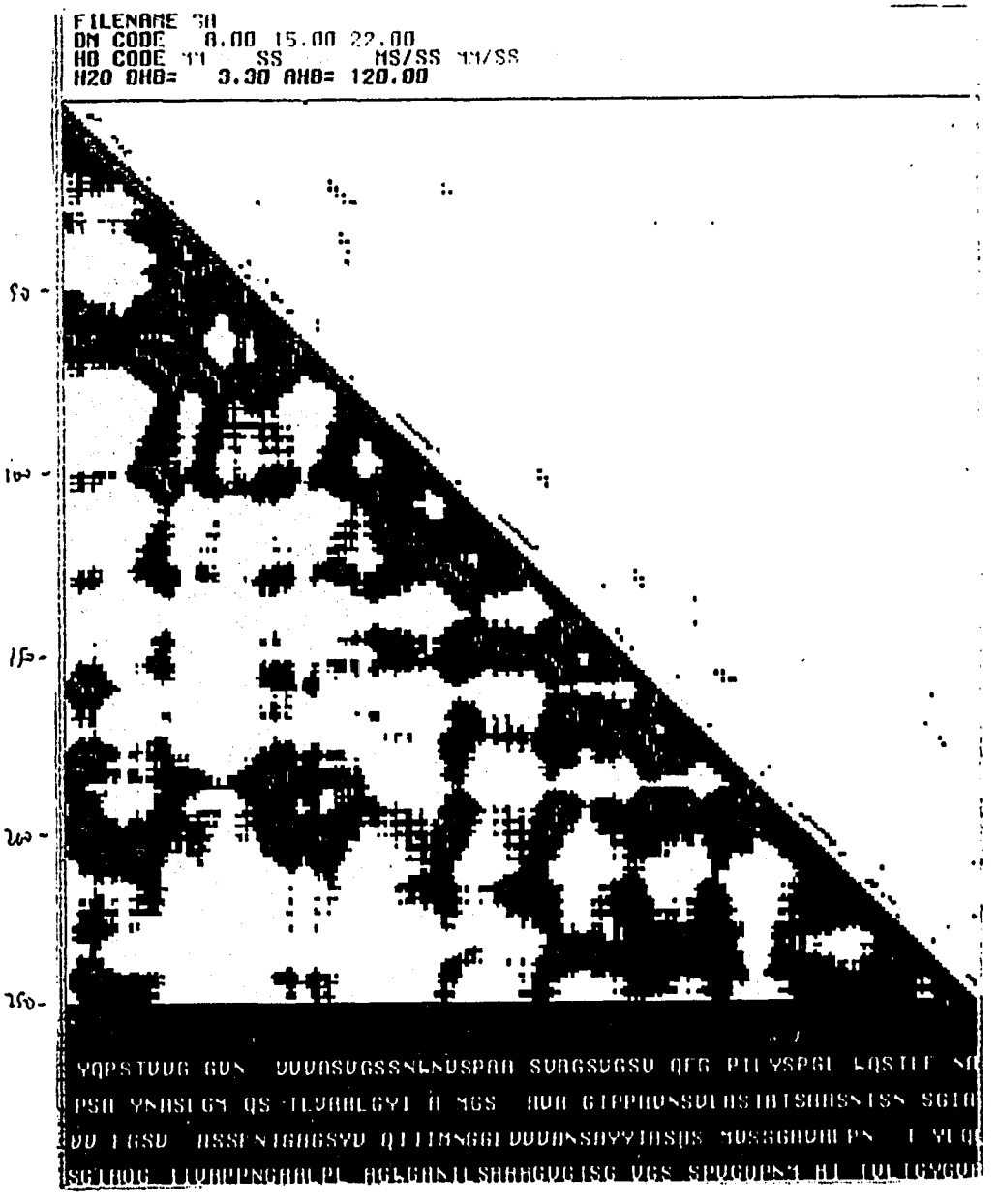


Figure 41a. DM/HBM plots of last 20ps average structure from SA (Ca^{2+} bound perturbation mutant) SBMD reduced dynamics simulation of subtilisin. There are only 258 residues in this structure; primary sequence numbers from native structure are shown in parentheses. DM is in bottom half of matrix, HBM is in top half of matrix and primary sequence is at bottom of plot. Contours of DM are 0-8Å(green), 8-15Å(light red), 15-22Å(dark red) and 22-30Å(orange). HB cutoffs are 3.3Å between heavy atoms and 120 degrees for A-H-D angle. HBM colors are MM(red), MS(yellow), SS(blue), MM/MS(orange), MS/SS(green), MM/SS(purple) and MM/MS/SS(brown). Primary sequence residues are colored by polarity: non-polar(white), apolar(yellow), acidic(red) and basic(blue).

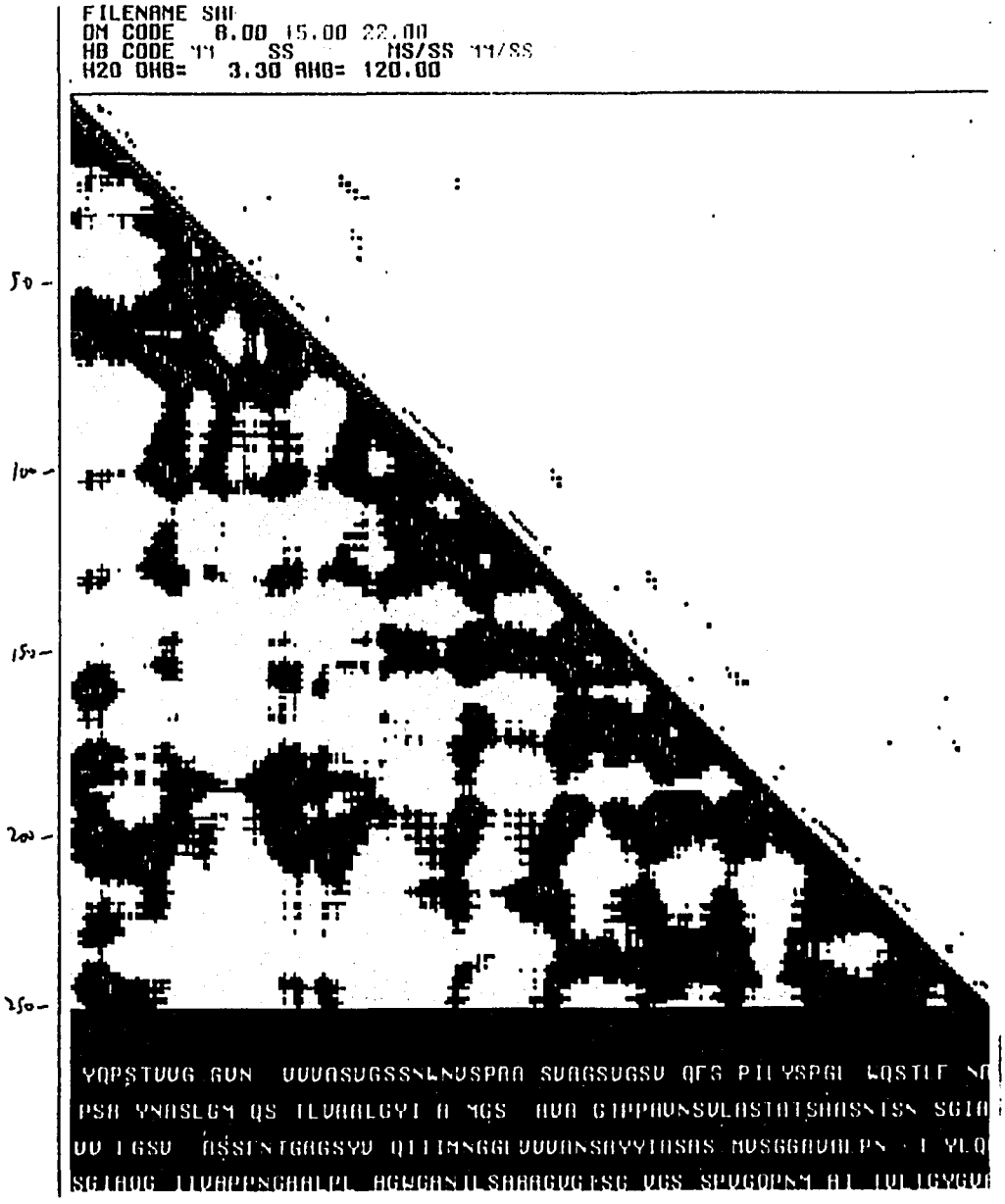


Figure 41b. DM/HBM plots of last 20ps average structure from SAF (apo perturbation mutant) SBMD reduced dynamics simulation of subtilisin. There are only 258 residues in this structure; primary sequence numbers from native structure are shown in parentheses. DM is in bottom half of matrix, HBM is in top half of matrix and primary sequence is at bottom of plot. Contours of DM are 0-8Å(green), 8-15Å(light red), 15-22Å(dark red) and 22-30Å(orange). HB cutoffs are 3.3Å between heavy atoms and 120 degrees for A-H-D angle. HBM colors are MM(red), MS(yellow), SS(blue), MM/MS(orange), MS/SS(green), MM/SS(purple) and MM/MS/SS(brown). Primary sequence residues are colored by polarity: non-polar(white), apolar(yellow), acidic(red) and basic(blue).

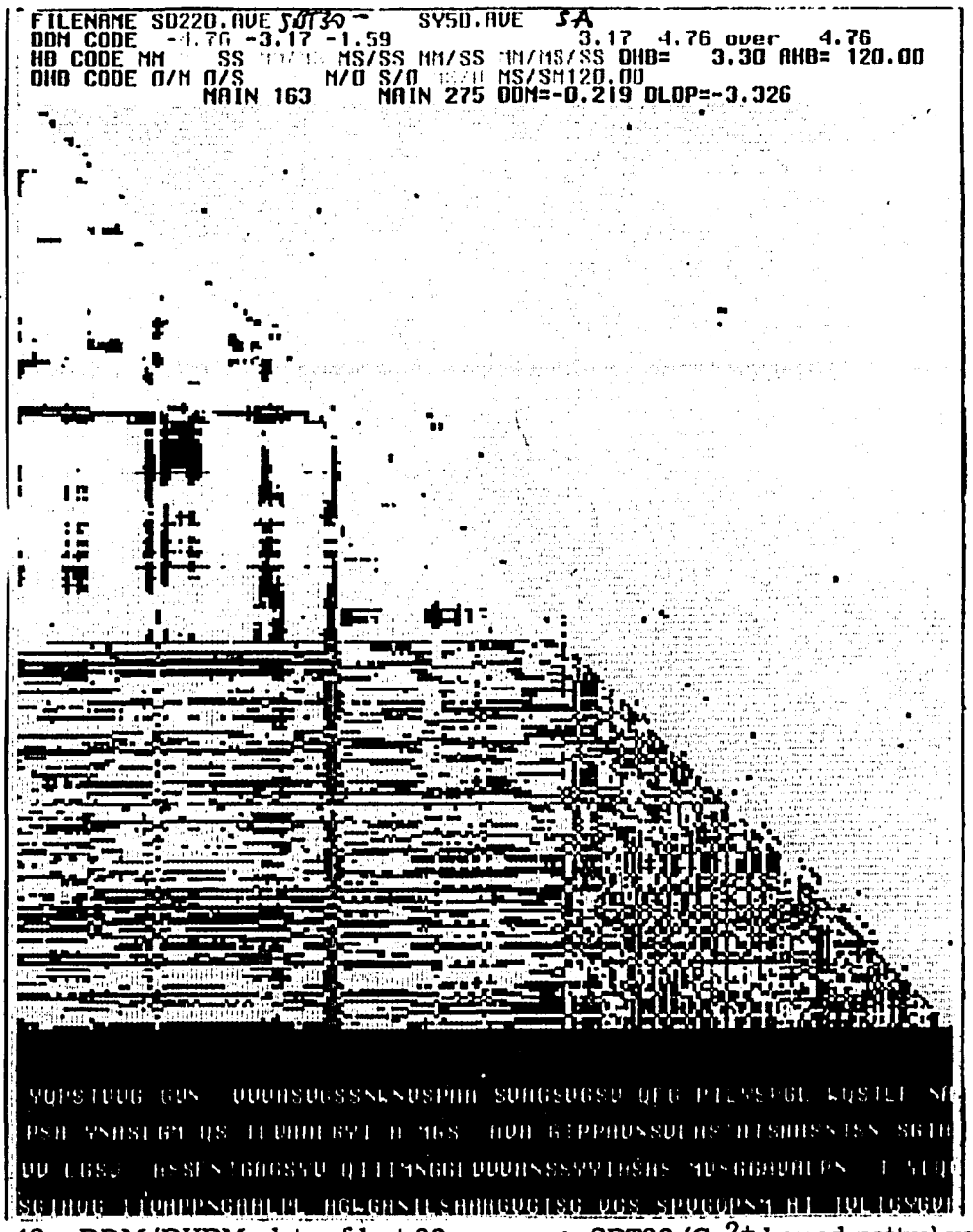


Figure 42a. DDM/DHBM plots of last 20ps average SBT30 (Ca^{2+} bound native) vs. last 20ps average SA (Ca^{2+} bound perturbation mutant) SBMD reduced dynamics subtilisin structures. There are only 258 residues in these structures; primary sequence numbers from native structure are shown in parentheses. DDM is in bottom half of matrix, DHBM is in top half of matrix and primary sequence is at bottom of plot. Contours of DDM (in angstroms) are multiples of standard deviation (1.59Å); less than-4.76(orange), -4.76 to -3.17(dark red), -3.17 to -1.59(red), -1.59 to 0.0(pink), 0.0 to 1.59(light blue), 1.59 to 3.17(blue), 3.17 to 4.76(dark blue) and greater than 4.76(black). HB cutoffs are 3.3Å between heavy atoms and 120 degrees for A-H-D angle. DHBM codes are O/M(red), O/S(blue), O/MS(yellow), M/O(dark red), S/O(dark blue), MS/O(orange) and MS/SM(purple) where O/X means gain and X/O means loss of hydrogen bond type X. Primary sequence residues are colored by polarity: non-polar(white), apolar(yellow), acidic(red) and basic(blue).

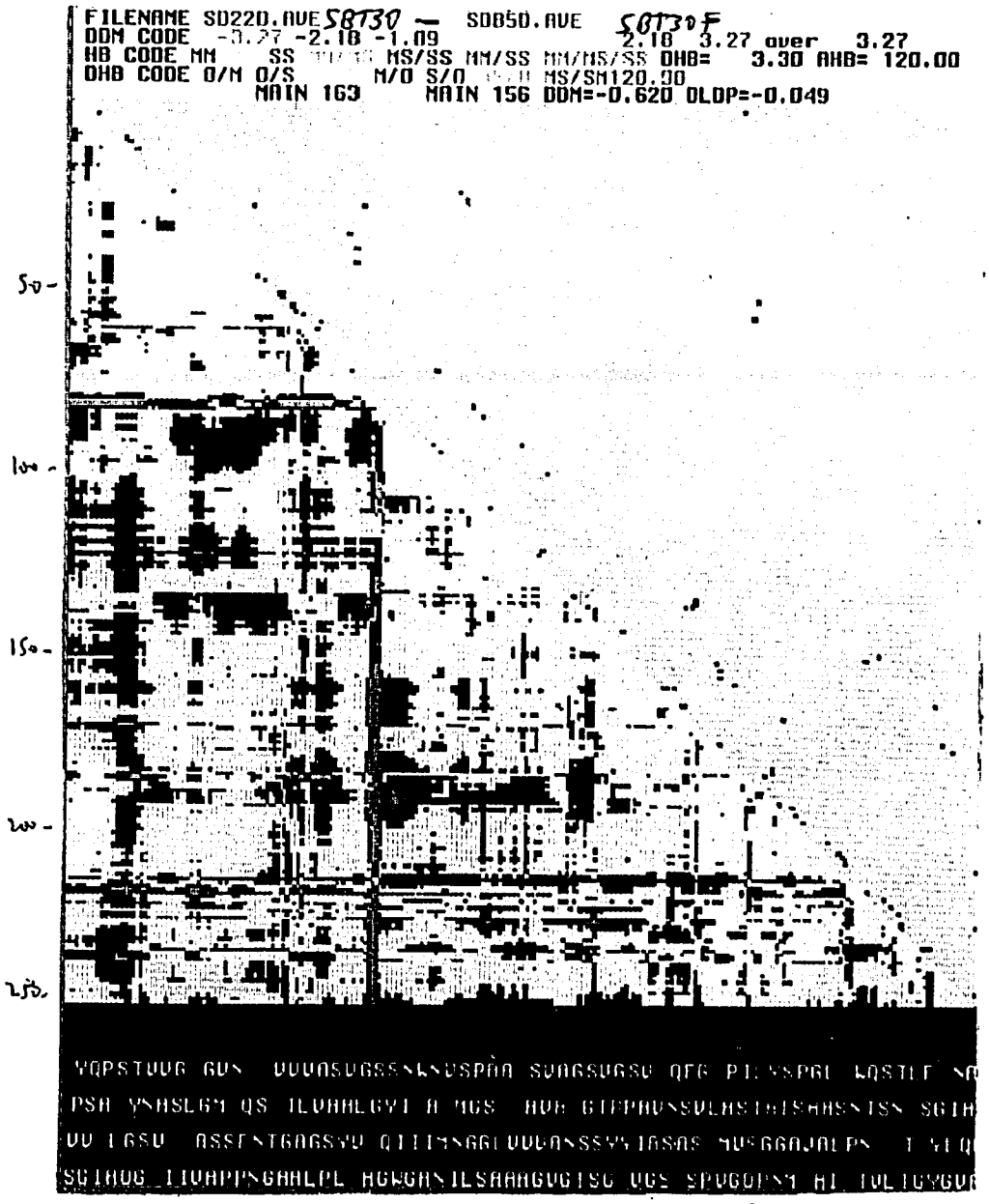


Figure 42b. DDM/DHBM plots of last 20ps average SBT30 (Ca^{2+} bound native) vs. last 20ps average SBT30F (apo native) SBMD reduced dynamics subtilisin structures. There are only 258 residues in these structures; primary sequence numbers from native structure are shown in parentheses. DDM is in bottom half of matrix, DHBM is in top half of matrix and primary sequence is at bottom of plot. Contours of DDM (in angstroms) are multiples of standard deviation (1.09Å); less than-3.27(orange), -3.27 to -2.18(dark red), -2.18 to -1.09(red), -1.09 to 0.0(pink), 0.0 to 1.09(light blue), 1.09 to 2.18(blue), 2.18 to 3.27(dark blue) and greater than 3.27(black). HB cutoffs are 3.3Å between heavy atoms and 120 degrees for A-H-D angle. DHBM codes are O/M(red), O/S(blue), O/MS(yellow), M/O(dark red), S/O(dark blue), MS/O(orange) and MS/SM(purple) where O/X means gain and X/O means loss of hydrogen bond type X. Primary sequence residues are colored by polarity: non-polar(white), apolar(yellow), acidic(red) and basic(blue).

D. The Ca^{2+} Binding Site: changes in loop-loop interactions

The structure of the Ca^{2+} binding sites from the native and mutant simulations of subtilisin are shown in the HBM and DM plots (of residues 145 to 210) displayed in Figure 43. The tertiary structures of the Ca^{2+} binding sites are very similar among all these structures (rms differences are ca. 1.0Å). However changes can been seen in the hydrogen bonding interactions among these structures as well as small changes in the conformation of loop 155 to 165. In the DM plot of the Ca^{2+} bound mutant the green line extending anti-parallel to the main diagonal, at ca. residue 165 in the upper left corner of the plot, represents a tighter turn and a more extended structure of this loop than observed in the other structures. In the apo native structure the HBM plot shows that the conformation of this loop has adopted a more compact structure; there are two red main chain-main chain hydrogen bonding interactions adjacent to the main diagonal around residue 165.

The hydrogen bonding interactions between loops 154-165 and 180-190 are adjacent to the Ca^{2+} binding loops 169-175 and 195-197. They also flank the binding site residues in the primary sequence of the protein. Thus it is possible to determine how changes in the conformations of these loops and the modes in which they interact are able to affect the structure and dynamics of the Ca^{2+} binding site, by following the pattern of hydrogen bonding.

In the upper part of the HBM plot of the native structure, in Figure 43, a blue side chain-side chain hydrogen bonding interaction is observed between residues 158 and 186. Extending to the main diagonal from this point, both down vertically and across horizontally, one can observe the local hydrogen bonding interactions among the residues in loops 155-165 and 180-190. In the native structure it is observed that there are very few hydrogen bonding interactions within the 155-165 loop whereas there are many interactions between the residues of the 180-190 loop. Hydrogen bond

interactions in the 180-190 loop, and not in the 155-165 loop, is the pattern observed for every structure except the apo native protein. In this apo structure there are numerous hydrogen bond interactions in the loop 155-165, and differences in the hydrogen bonding interactions between this loop and 180-190.

In the Ca^{2+} bound and apo native and mutant structures there are four conserved hydrogen bond interactions involving the loop containing residue 163. These are the main chain-main chain hydrogen bonds between the A153O and G166N, V192O and T164N, G166O and K170N, and P168O and Y171N.

In the native structures, both Ca^{2+} bound and apo, there is a hydrogen bond interaction between the side chain hydroxyl of T164 and the carbonyl of G157 while in the mutant structures the side chain hydroxyl of T164 is hydrogen bonded to the carbonyl of S162. In the apo native structure three additional hydrogen bond interactions, not observed in the mutant or Ca^{2+} bound native structures, are found between residues 158 and 162; between residues T158 and G160, S159 and S162, and between G160 and S162.

A salient structural difference between the mutant and native structures, immediately adjacent to residue 163, is the hydrogen bond between residues 162 and 164 that forms in the mutants. In the absence of Ca^{2+} the native structure forms several hydrogen bond interactions between residues 158 to 164 that are absent in the mutant, presumably due to the conformational change at residues 162 to 164 resulting from the mutation. This difference in the structure and interactions involving residues 158 to 164 leads to a different conformational response to the absence of Ca^{2+} in the interactions of loop 154-165 with loop 180-190: In the absence of Ca^{2+} the apo native structure forms several additional hydrogen bonds between T158 and residues 188 and 190, while in the mutant structures (apo and Ca^{2+} bound) the only hydrogen bond between these residues is the one between the side chains of T158 and R186 which is present in all native and mutant structures. (In the Ca^{2+} bound mutant the distance of

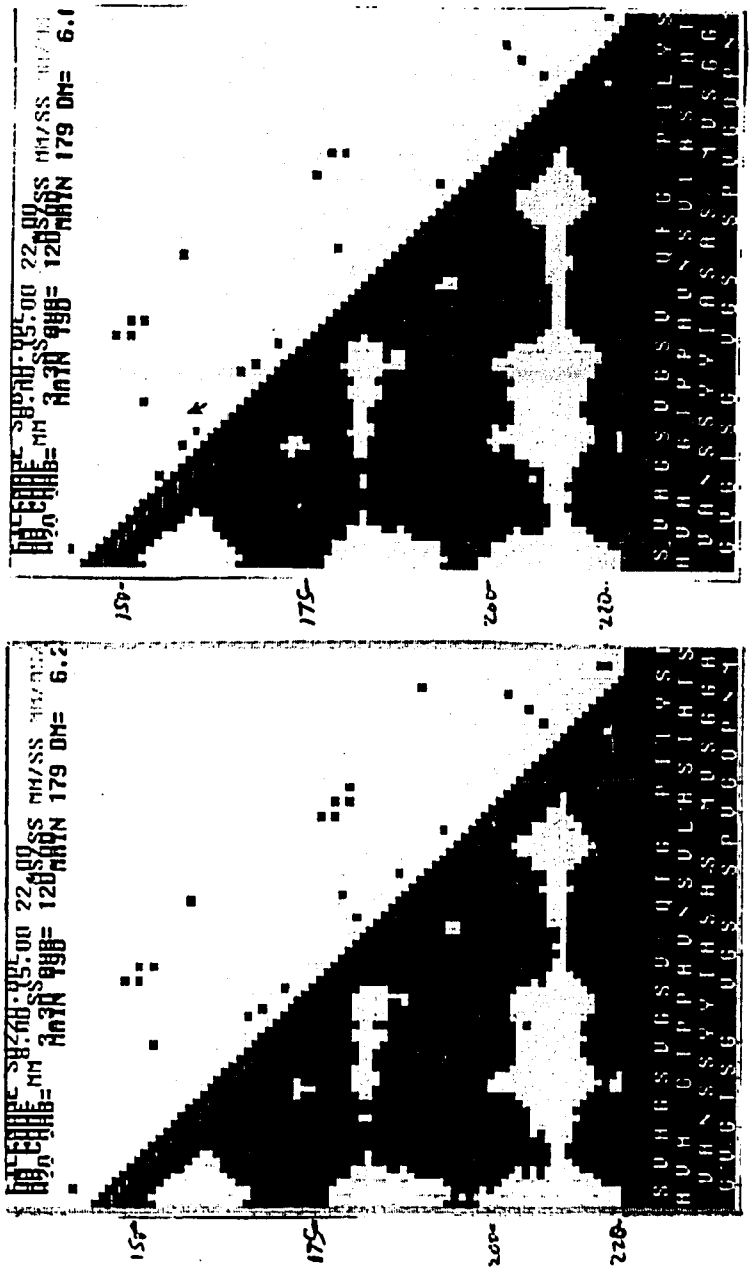


Figure 43a-b. DM/HBM plots of residues 145 to 220 from last 20ps average structures from SBT30 (Ca^{2+} bound native; left plot) and SBT30F (apo native; right plot) SBMD reduced dynamics simulations of subtilisin. There are only 258 residues in these structures however the axes correspond to primary sequence numbers from native structure. DM is in bottom half of matrix, HBM is in top half of matrix and primary sequence is at bottom of plot. Contours of DM are 0-8Å(green), 8-15Å(light red), 15-22Å(dark red) and 22-30Å(orange). HB cutoffs are 3.3Å between heavy atoms and 120 degrees for A-H-D angle. HBM colors are MM(red), MS(yellow), SS(blue), MM/MS(orange), MS/SS(green), MM/SS(purple) and MM/MS/SS(brown). Primary sequence residues are colored by polarity: non-polar(white), apolar(yellow), acidic(red) and basic(blue).

Figure 43c-d. DM/HBM plots of residues 145 to 220 from last 20ps average structures from SA (Ca^{2+} bound perturbation mutant; left plot) and SAF (apo perturbation mutant; right plot) SBMD reduced dynamics simulations of subtilisin. There are only 258 residues in these structures however the axes correspond to primary sequence numbers from native structure. DM is in bottom half of matrix, HBM is in top half of matrix and primary sequence is at bottom of plot. Contours of DM are 0-8Å(green), 8-15Å(light red), 15-22Å(dark red) and 22-30Å(orange). HB cutoffs are 3.3Å between heavy atoms and 120 degrees for A-H-D angle. HBM colors are MM(red), MS(yellow), SS(blue), MM/MS(orange), MS/SS(green), MM/SS(purple) and MM/MS/SS(brown). Primary sequence residues are colored by polarity: non-polar(white), apolar(yellow), acidic(red) and basic(blue).

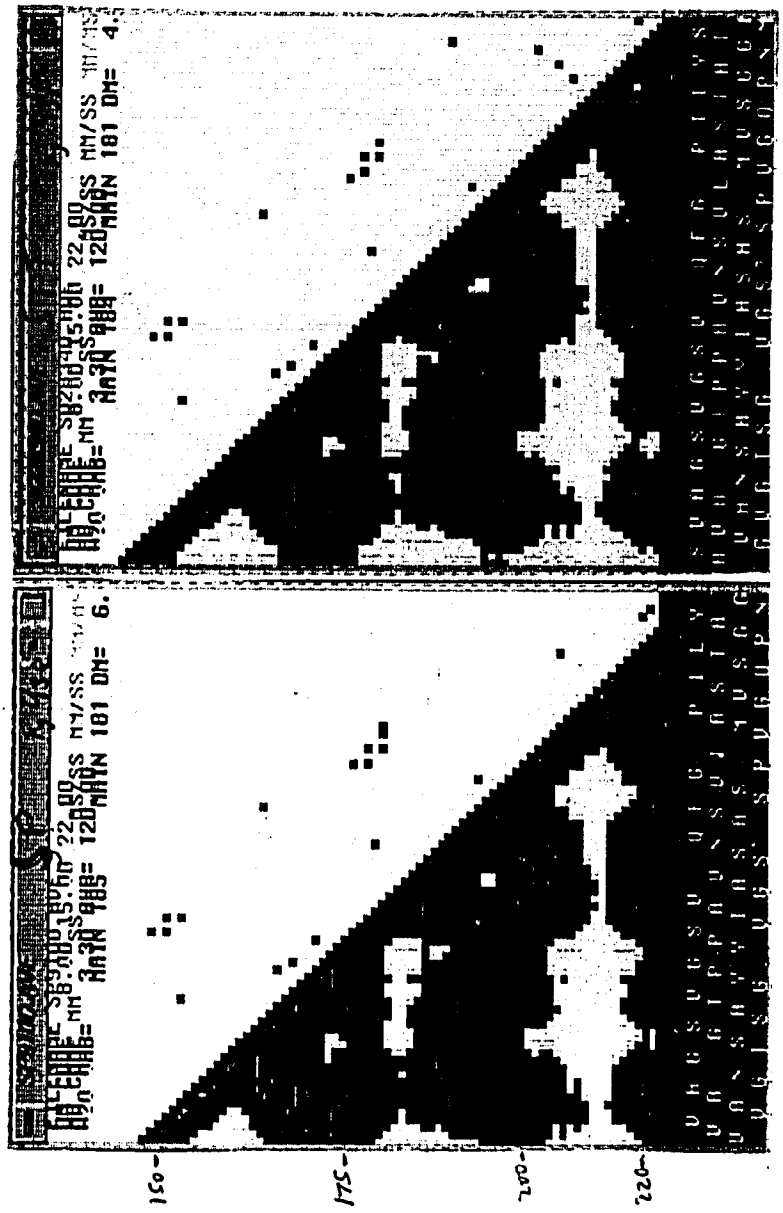


Figure 43e-f. DM/HBM plots of residues 145 to 220 from last 20ps average structures from SP (Ca^{2+} bound sudden deletion mutant; left plot) and SPF (apo sudden deletion mutant; right plot) SBMD reduced dynamics simulations of subtilisin. There are only 258 residues in these structures however the axes correspond to primary sequence numbers from native structure. DM is in bottom half of matrix, HBM is in top half of matrix and primary sequence is at bottom of plot. Contours of DM are 0-8Å(green), 8-15Å(light red), 15-22Å(dark red) and 22-30Å(orange). HB cutoffs are 3.3Å between heavy atoms and 120 degrees for A-H-D angle. HBM colors are MM(red), MS(yellow), SS(blue), MM/MS(orange), MS/SS(green), MM/SS(purple) and MM/MS/SS(brown). Primary sequence residues are colored by polarity: non-polar(white), apolar(yellow), acidic(red) and basic(blue).

this conserved hydrogen bond has extended to 3.35Å). This difference has a direct effect on the interactions between the two loops in the Ca^{2+} binding site and the molecular dynamics properties expressed by the binding site.

The distances of oxygen ligands in the Ca^{2+} binding sites for all the structures is shown in Table 8. The coordination distances in the mutant structures are shorter than those in the native structure. However, compared to the average coordination distances in the solvated simulation (shown in Table 5) this difference is not significant.

To evaluate these differences for the binding of Ca^{2+}, a calcium ion was replaced into the apo native and apo-perturbation structures, and the position of the ion was optimized with the protein and solvent fixed. The results show that the Ca^{2+} replaced into the apo mutant structure binds with a coordination to only three solvent molecules, whereas the Ca^{2+} replaced into the apo native structure was coordinated to six solvent molecules. Thus the binding site of the apo native structure appears more solvent accessible than in the mutant, suggesting a role of serine 163 in maintaining

Table 8. Distances of oxygen ligands from Ca^{2+} in the mutant subtilisin simulations and the starting structure. Standard errors are in parentheses.

LIGAND	SBT30	SA	SP
PRO 172 O	2.41	2.41 (0.09)	2.42 (0.08)
VAL 174 O	2.68	2.48 (0.10)	2.47 (0.09)
ASP 197 OD1	2.78	2.56 (0.08)	2.56 (0.09)
ASP 197 OD2	2.50	2.50 (0.06)	2.49 (0.08)
TIPS 42 OH2	2.28	2.39 (0.06)	2.38 (0.06)
TIPS 55 OH2	2.67	2.42 (0.08)	2.46 (0.11)
TIPS 74 OH2	2.43	2.37 (0.06)	2.39 (0.07)
SOLV 703 OH2	2.46	2.42 (0.09)	2.40 (0.08)

SBT30 = 440ps structure from SBMD simulation
SA = S163A perturbation simulation
SP = S163A sudden deletion simulation
TIPS = Crystal water
SOLV = Water added during solvation procedure

the conformation of the Ca^{2+} binding site in the absence of Ca^{2+}. Once Ca^{2+} binds, the ion is the dominant organizational force in the structure of the binding site. While the observed differences in coordination of the mutant and native Ca^{2+} bound structures are not statistically significant, there are large differences observed in the dynamics properties of these structures.

E. Effects of mutations on the dynamic properties of the protein

The alpha carbon CDM plots of the native and mutant proteins from the apo and Ca^{2+} bound simulations are shown in Figure 44. These plots are of protein structures, from reduced dynamics simulations. As a result of the removal of 17 residues from the protein reservior region the axes of the CDM plots encompass only 258 residues. The deletion of 12 residues after 75 causes a shift in the numbering (i.e. residue 150 on the plot is residue 162 in the primary sequence) and there is a discontinuity in the plot at residue 201 which represents the deletion of residues 211 to 214 from the reservior. The highest correlation is along the main diagonal and shown in purple. The yellow represents negatively correlated dynamical motions and the blue represents little correlation.

The structural changes observed in the DDM plots (Figure 42) show that the mutation affects the conformation of the carboxy terminal domain and here the dynamics properties from the mutant simulations show that there is a correlation between the structural changes in the carboxy domains and dynamical interactions between the Ca^{2+} binding site and these domains.

In the CDM plot of the Ca^{2+} bound native structure, the Ca^{2+} binding site (residues 150 to 200 in the CDM plot) shows highly correlated dynamical behavior. There is very high correlation in the molecular dynamics of the loops in the Ca^{2+} binding site as well as to the loop containing residues 154-165, shown as residues 143-154 in the CDM plot. There is also a high correlation between the loop containing

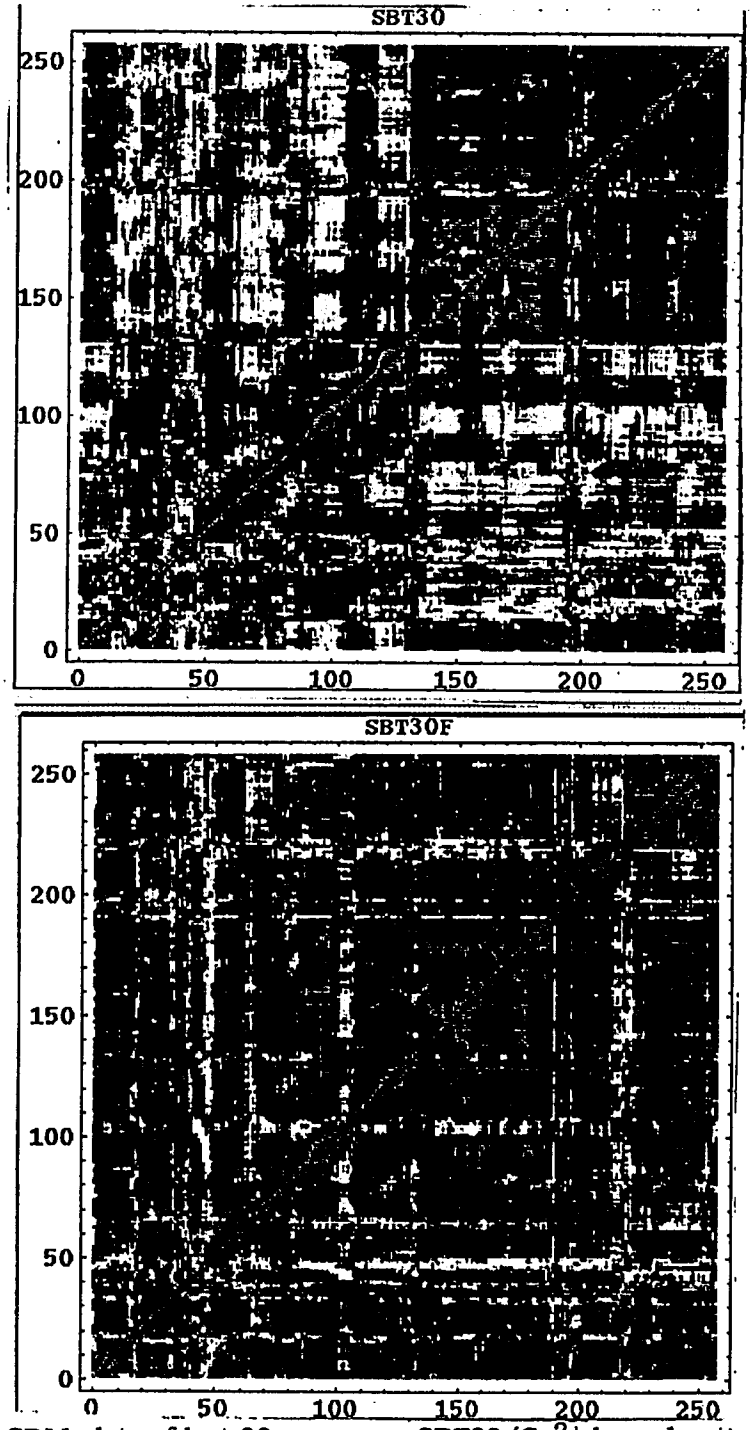


Figure 44a-b. CDM plots of last 20ps average SBT30 (Ca^{2+} bound native, top plot) and SBT30F (apo native, bottom plot) SBMD reduced dynamics subtilisin structures. There are only 258 residues in these structures; primary sequence numbers from native structure are shown in parentheses. Matrices are square symmetric and represent normalized correlation coefficients (-1 to +1) of alpha carbon positional displacements. Correlations are color coded as; very high (greater than 0.8, red), high (0.5 to 0.7, purple), low positive (0.0 to 0.5, blue) and low negative (-0.5 to 0.0, yellow).

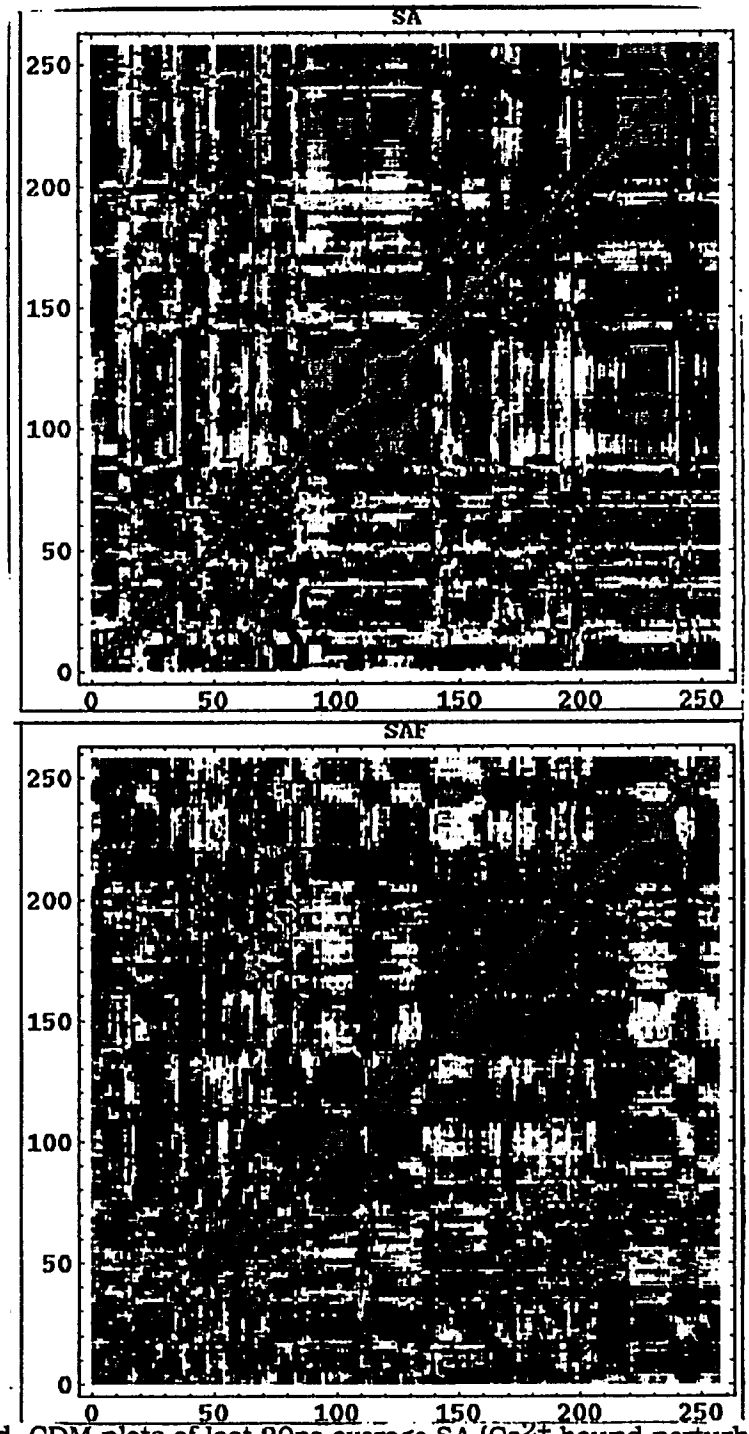


Figure 44c-d. CDM plots of last 20ps average SA (Ca^{2+} bound perturbation mutant, top plot) and SAF (apo perturbation mutant, bottom plot) SBMD reduced dynamics subtilisin structures. There are only 258 residues in these structures; primary sequence numbers from native structure are shown in parentheses. Matrices are square symmetric and represent normalized correlation coefficients (-1 to +1) of alpha carbon positional displacements. Correlations are color coded as; very high (greater than 0.8, red), high (0.5 to 0.7, purple), low positive (0.0 to 0.5, blue) and low negative (-0.5 to 0.0, yellow).

residues 190-195 and the segment around residue 240. Upon mutation of Ser163, these dynamical interactions are lost along with the conformational changes in the carboxy terminal domains. In the Ca^{2+} bound mutant the correlation between the dynamics of the Ca^{2+} binding site and the structure around residue 240 is lost and replaced by correlations between these two structures and the substrate binding site. There is also a large increase in positively correlated dynamics within the carboxy terminal domains. The effect of the mutation is to disrupt the dynamics of the Ca^{2+} binding site and its interactions with other tertiary structures in the protein.

The correlated dynamics of the loops 155-165 and 180-190 with the Ca^{2+} binding site is also greatly reduced in the apo simulations of the mutant and native structures. However, in other tertiary structures of the protein, the perturbation in the dynamics properties observed in the apo simulations is different from the perturbation observed in the mutation simulations. In the apo native structure there is a large attenuation throughout the protein of negatively correlated motions and an increase in positive correlation in the dynamics interactions between the Ca^{2+} binding site and the substrate binding site. In the apo mutant structure the only large change in the dynamics properties from the native structure are just the loss of positively correlated motions within the Ca^{2+} binding site.

The mutation of Ser163 results in conformational changes in the loop containing residues 154-165, and in the absence of Ca^{2+}, affects the interactions of this loop with the loop containing residues 180-190. These structures are adjacent to the Ca^{2+} binding site and the entropic cost of binding Ca^{2+} may be increased by the perturbation in the dynamics of this structure resulting from the mutation. These findings suggest a mechanism by which the mutation may result in a lower Ca^{2+} affinity at this binding site.

Along with the shorter coordination distances observed in the mutant structures (in Table 7) there were large changes in the interactions between residues 163,

Table 9. Distances between selected atoms of residues S163, K170 and E195 in mutant and native subtilisins.

ATOMS	SBT30	SBT30F	SA	SAF
K170NZ-E195CD	5.71 (0.33)	5.74 (0.31)	6.97 (0.50)	6.19 (0.28)
S163CB-E195CD	4.47 (0.21)	6.01 (0.35)	6.28 (0.46)	4.95 (0.23)
S163CB-K170NZ	6.22 (0.29)	5.32 (0.33)	6.57 (0.45)	6.27 (0.43)
K170CA-E195CA	5.13 (0.18)	5.66 (0.31)	5.58 (0.23)	5.50 (0.27)
S163CB-E195CA	6.37 (0.24)	8.84 (0.38)	7.11 (0.23)	6.81 (0.18)
S163CB-K170CA	8.47 (0.25)	9.40 (0.26)	9.04 (0.32)	8.62 (0.29)

SBT30 = Solvated SBMD simulation of native Ca^{2+} bound subtilisin
SBT30F = SBT30 with Ca^{2+} removed
SA = Simulation of S163A perturbation mutant of subtilisin
SAF = SA with Ca^{2+} removed

K170 and E195 resulting from the absence of Ca^{2+} and the mutations. Table 9 shows the distances between selected atoms from these residues observed in the mutant and apo simulations and Table 10 shows the correlation in dynamics among these residues.

While the hydrogen bonding distances between the side chain of S163 and the side chains of K170 and E195, observed in the xray structure, were not observed in the simulations of the native or mutant structures, nevertheless there was significant differences in the interactions between these residues observed in these simulations. In Table 10 in can be observed that there was a very high positive correlation between the motions of the three side chains (with an average correlation of 0.71). In the apo native structure the correlations between these side chains are still large, decreasing to an average of 0.51, while in the mutant structure the correlation among the motions of these side chains is significantly reduced (to an average of 0.16). In Table 9 it is observed that there is an increase in the distance between the side chains of these residues, from the native structures to the Ca^{2+} bound mutant, along with an increase in the fluctuations in position of these atoms. These results show that a large positive

Table 10. Dynamics covariances between selected atoms in residues S163, K170 and E195 in mutant and native subtilisins.

ATOMS	SBT30	SBT30F	SA	SAF
K170NZ-E195CD	0.69	0.44	0.24	0.47
S163CB-E195CD	0.72	0.46	0.26	0.62
S163CB-K170NZ	0.71	0.62	0.02	0.53
K170CA-E195CA	0.18	0.26	-0.15	-0.22
S163CA-E195CA	-0.15	0.08	-0.07	-0.02
S163CA-K170CA	0.14	0.22	0.02	0.09

SBT30 = Solvated SBMD simulation of native Ca^{2+} bound subtilisin
SBT30F = SBT30 with Ca^{2+} removed
SA = Simulation of S163A perturbation mutant of subtilisin
SAF = SA with Ca^{2+} removed

interaction between the side chains of these residues is maintained in the native structures even in the absence of close hydrogen bonding.

In Table 10 a small positive correlation can be observed in the dynamics of the alpha carbons of the Ca^{2+} bound and apo native structures between residue K170 and residues S163 and E195 (averaging 0.16 in the Ca^{2+} bound and 0.24 in the apo native structure). In the mutant structures there is no correlation (ca. 0.0) between the alpha carbons of residue 163 and the two residues of the Ca^{2+} binding site (K170 and E195). There is a negative correlation between residues K170 and E195 in the Ca^{2+} bound (-0.15) and the apo (-0.22) mutant structures. The dynamics interactions between the loops containing these side chains (also observed in the CDM plots) are positively correlated in the native structures and this changes into negatively correlated motions in the mutant structures.

V. DISCUSSION

A. Accuracy of Stochastic Boundary Method in Molecular Dynamics Simulations

PG3 and Erabutoxin Studies

The analysis of the structure and dynamics of the cyclo(L-pro-gly)3-Ca^{2+} complex resulting from three different boundary conditions show that there is no significant difference between the results of the stochastic boundary (SBMD) method and the standard periodic boundary (BOX) method. In contrast, differences in both the structure and dynamics of the peptide-ion complex, compared to the previous two simulations, were observed with the harmonic boundary (HAR) method. Thus while the HAR method is easily constructed its results are not as accurate as those from the SBMD method when the periodic boundary method is used as a standard of comparison.

The comparison of the unbounded solvated simulation of erabutoxin with a simulation using SBMD boundary conditions showed that there was no significant effect on the protein from the SBMD boundary conditions.

The SBMD method is an accurate procedure for simulating solvated macromolecular structures. The SBMD reduced dynamics can study molecular processes that occur in systems and time scales where a full molecular dynamics simulation is not feasible, which makes it applicable to a wider range of problems than the BOX method.

B. Importance of Solvent in Molecular Dynamics of Proteins

Role of Solvent in the Proposed Mechanism of Erabutoxin Binding

A model for the mechanism of erabutoxin b binding to the acetylcholine receptor has been proposed by Dr. Low [Low, 1986, 1987]. This mechanism requires a large structural change in the protein leading to side-chain stacking interactions between a tryptophan (residue 29) in the snake venom and one in the receptor. This mechanism hypothesizes a high degree of mobility in the segment of erabutoxin containing residues 45-55. These residues contact tryptophan 29 in the xray structure. The hypothesized mechanism of binding involves a movement of the segment away from tryptophan 29, followed by binding to a tryptophan residue in the receptor and reassociation of this protein segment (residues 45-55) shielding the two tryptophans from solvent. The plots of backbone and sidechain mobility calculated from both the EBS and EBU simulations of solvated erabutoxin b are displayed in Figure 22 and, in agreement with the xray data (Figure 1b, Low, 1986), suggest that this segment is highly mobile. The plots from both simulations are very similar to the xray data in both the pattern and magnitude of mobility of the residues. Thus the results obtained from the molecular dynamics simulations support the proposed mechanism of erabutoxin b binding to the acetylcholine receptor. On the methodological side, the similarity between EBS and EBU results indicate that there is no significant effect of the SBMD boundary condition on the protein dynamics when compared to the unbounded simulation.

Comparison of Solvated Subtilisin Simulations

From the simulation of solvated low affinity Ca^{2+} binding site of subtilisin, using the SBMD boundary conditions, conformational changes were observed, between

the N-terminal domain (in the buffer) and the rest of the protein, which were associated with changes in solvent networks observed in the xray structure. Some of these conformational changes are observed (i.e. the domain interface at the helix containing residues 64-73) in the absence of solvent. The interdomain separation between the N-terminal domain and the rest of the protein in the solvated simulation result from the free interaction of the two carboxy domains with solvent while the N-terminal domain was constrained by the boundary conditions.

While the coordination structure changes only slightly in the *in vacuo* simulation, in the solvated simulation there are conformational changes in the Ca^{2+} binding site. These changes were also observed in a full crystal simulation of subtilisin [Heiner, 1992] and thus are not the result of the SBMD boundary conditions used in this simulation. These conformational changes are not observed in the *in vacuo* simulation and these results emphasize that solvent is essential to the molecular dynamics properties of this protein.

C. Role of Serine 163 in Structure and Dynamics of Ca^{2+} Site B in Subtilisin

Comparison of Effects of Serine Mutation in Apo and Calcium Bound Structures

In the absence of the interactions between the side chains of residue S163 with residues K170 and E195 the dynamics of the loops in the Ca^{2+} binding site are affected. Residue S163 in subtilisin is necessary to maintain the highly correlated dynamics properties between the two loops of the Ca^{2+} binding site B (residues 169-175 and 195-197).

In simulations of the S163A mutant a hydrogen bond interaction between the side chain of residue T164 and the backbone carbonyl of S162 was observed to form. The conformation of the loop 155-165, containing the mutated residue 163, was

significantly affected by this hydrogen bond as was the interactions between this loop and residue S191 (which is adjacent to loop 195-197 of the Ca^{2+} binding site).

In simulations in the absence of Ca^{2+} hydrogen bonding interactions between residues 154-158 and residues 186-192 were observed in the native but not the mutant structures. The structure of the loop 155-165, and its interactions with loop 180-190, differed significantly in the native Ca^{2+} bound and apo structures. In the mutant Ca^{2+} bound and apo structures there was no significant differences in the interactions between these two loops.

In the Ca^{2+} bound mutant the dynamics of the Ca^{2+} binding site was significantly different from the dynamics of the native binding site. The high positively correlated dynamics motions of the Ca^{2+} binding site actually became highly negatively correlated. The mutation also resulted in large conformational changes throughout the carboxy terminal domain of the protein as well as large perturbations in the dynamics of this domain.

Analysis of both the structural differences and differences in dynamics properties expressed by the Ca^{2+} bound vs. apo simulations of the native and mutant subtilisins suggest that the S163A mutant would express a higher affinity for Ca^{2+} than the native protein (see next section: VI. Concluding Remarks).

Finally, the absence of Ca^{2+}, both in the native and mutant structures, was observed to affect the structure of the substrate binding site as well as the dynamics of the interactions between the Ca^{2+} binding site and the substrate binding site, suggesting that Ca^{2+} may modulate substrate binding to the protein.

VI. CONCLUDING REMARKS

Calcium Ion Affinity of the S163A Mutant Subtilisin

From the CDM plots from the reduced dynamics simulations of the Ca^{2+} bound and apo native and mutant subtilisin structures shown in Figure 44 one can observe significant differences in the correlated dynamics motions of the low afffinity Ca^{2+} binding site from the simulations of the native protein (Ca^{2+} bound vs. apo). These large differences are not observed in the results from the simulations of the mutant structures. Using a normal mode approximation, it is possible to estimate the difference in configurational entropy between the Ca^{2+} bound and apo structures [Karplus, 1981] from the molecular dynamics trajectories. While the entropic difference is a function of the changes that occur throughout the entire molecule, the large difference observed in the Ca^{2+} binding site between the Ca^{2+} bound and the apo native structures strongly indicate that the difference in entropy in the mutant structures is much smaller than in the native protein. Thus, the estimated entropic cost of binding Ca^{2+} is larger in the native than the mutant.

The ligand distances obtained from the simulations of the mutant subtilisin structures are shown in Table 8 and those obtained from the simulations of the native subtilisin structure are shown in Table 5. The coordination distances of the ligands taken as a group are not significantly different. The only exception is the main chain carbonyl of Val 174; the coordination distance to this ligand is significantly shorter in the mutant than the native structure.

The coordination distances are largely the result of the Ca^{2+} parameters used which yield optimal Ca^{2+} coordination distances of ca. 2.4Å. The coordination distances, overall, are not significantly different between the native and mutant structures. The smaller coordination distance of Val 174 in the mutant compared to the

native structure indicates that the enthalpic difference in the Ca^{2+} bound vs. apo structures may be larger in the mutant than in the native protein. It is noteworthy that the coordination structure of the 440ps conformation from the SBT30 native subtilisin simulation (used as the starting structure for the simulations of the mutant subtilisin) has larger coordination distances than the averages in the SBT30 simulation (see Table 8).

The structural and dynamics results from these simulations indicate that both the enthalpic and the entropic changes involved in the binding of Ca^{2+} to the mutant structure result in a higher affinity than in the native protein.

Structure of the Calcium Bound Low Affinity Site B in Subtilisin

The low affinity binding site in subtilisin is known to bind several different ions besides Ca^{2+} (notably sodium and potassium ions). The affinity of this site for Ca^{2+} is only ca. 10^{-4}M. The Ca^{2+} in the low affinity binding site of the xray structure of subtilisin, used as the starting conformation for the simulations in this thesis, has an occupancy of only ca. 50%. Thus, the coordination structure of the binding site, obtained from the xray crystallographic study, may represent a mixture of states in which several different ions may be bound. This is an additional complication for the interpretation of the changes observed in the coordination structure of the Ca^{2+} binding site during the simulation of the native structure (the SBT30 simulation).

The results of the simulation in this study of the solvated Ca^{2+} bound low affinity site B in subtilisin (the SBT30 simulation) agree in many aspects with the full crystal simulation reported by Heiner, et. al. [Heiner, 1992]. Specifically, both simulations resulted in shorter, more optimal coordination distances than were observed in the xray structure. Also, there was a significant shift in the tertiary structure of the Ca^{2+} binding site relative to other structures in the protein. Both of these observations, along with the observation that the Ca^{2+} binding site is not fully

occupied in the xray structure, suggest that the conformational changes observed in the Ca^{2+} binding site during the simulations may represent an equilibration toward the actual Ca^{2+} bound state in solution.

A final point to emphasize is that the low affinity Ca^{2+} binding site B in subtilisin is a highly unusual structure when compared to other, highly structurally characterized, Ca^{2+} binding sites in proteins. Unlike site B, most Ca^{2+} binding sites in proteins consist of ligands from residues closely spaced in the primary sequence. An example is the family of Ca^{2+} binding proteins containing EF hands (calmodulin, troponin C, calbindin, etc.) in which the coordinating ligands are from a single 12 residue segment of the primary sequence. Also, in the high affinity binding site A in subtilisin one segment of the protein (residues 76 to 81) contributes four coordinating ligands. In these sites it is believed that the flexibility of the binding site is very low.

While one may often think of Ca^{2+} binding sites as having a low flexibility, the low affinity binding site in subtilisin is structurally unusual and, thus, the flexibility of the protein segments in this site, as well as its dynamics properties, may be significantly different from other binding sites that have been characterized.

It has been proposed that the formation of the native structure represents a significant barrier (due to electrostatic repulsion) to the folding of the protein [Bryan, 1992]. This suggests that the conformational change associated with ion binding may be much larger than assumed from studies of other Ca^{2+} binding sites. Thus the behavior of the Ca^{2+} binding site observed during these simulations may be accurately representing significant properties of this structure.

As van Gunsteren has stated [Van Gunsteren, 1988], until the recent developments in computational chemistry, theoretical methods had not played a significant role in biochemistry. As a result of the need for basic knowledge of protein chemistry and the concurrent advances in the field of computational chemistry, protein simulations should become an important method when applied to

structure/activity relationships. These simulations complement the experimental work on protein structure and this interaction should result in progress in both these areas.

VII. REFERENCES

Alder, B. J., and Wainwright, T. E.; "Studies in Molecular Dynamics. I. General Method"; J. Chem. Phys., 31[1959],459-466.

Axelsen, P. H., Haydock, C., and Prendergast, F. G.; "Molecular Dynamics of Tryptophan in Ribonuclease-T1"; Biophys. J., 54[1988]249-258.

Babu, Y. S., Bugg, C. E., and Cook, W. J.; "Crystal Structure of Calmodulin"; in Calcium-binding Proteins in Health and Disease, pp. 305-321; ed. Norman, A. W., et. al., pub. Academic Press, New York, 1987.

Bajorath, J., Raghunathan, S., Hinrichs, W., and Saenger, W; "Long-range structural changes in proteinase K triggered by calcium ion removal"; Nature, 337[1989]481-484.

Berendsen, H. J. C.; "Molecular dynamics studies of proteins and nucleic acids"; Curr. Opin. Struc. Biol., 1[1991]191-195].

Blackburn, S.; "Subtilisin"; chap. 5 in Enzyme Structure and Function, by Blackburn, S.; pub. Marcel Dekker, Inc., New York, 1976.

Blumenthal, D. K., and Stull, J. T.; "Activation of Skeletal Muscle Myosin Light Chain Kinase by Calcium(2+) and Calmodulin"; Biochem., 19[1980]5608-5614.

Bott, R. R., Ultsch, M., Kossiakoff, A., Graycar, T., Katz, B., and Power, S.; "The three-dimensional structure of Bacillus amyloliquifaciens subtilisin at 1.8Å and an analysis of the structural consequences of peroxide inactivation"; J. Biol. Chem., 263[1988]7895-7906.

Bourne, P. E., Sato, A., Corfield, P. W. R., Rosen, L. S., Birken, S., and Low, B. W.; "Erabutoxin b. Initial protein refinement and sequence analysis at 0.140-nm resolution"; Eur. J. Biochem., 153[1985]521-527.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M.; "CHARMm: A Program for Macromolecular Energy, Minimization and Dynamics Calculations"; J. Comp. Chem., 4[1983b]187-217.

Brooks, C. L. III, Brunger, A., and Karplus, M.; "Active Site Dynamics in Protein Molecules: A Stochastic Boundary Molecular-Dynamics Approach"; Biopolymers, 24[1985]843-865.

Brooks, C. L. III, and Karplus, M.; "Deformable stochastic boundaries in molecular dynamics"; J. Chem. Phys., 79[1983a]6312-6325.

Brooks, C. L. III, and Karplus, M.; "Theoretical approaches to solvation of biopolymers"; Meth. Enzymol., 127[1986]369-400.

Brooks, C. L. III, Karplus, M., and Pettitt, B. M.; "Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics"; John Wiley and Sons, New York, 1988.

Brostrom, M. A., Brostrom, C. O., Breckenridge, B. McL., and Wolff, D. J.; "Calcium-Dependent Regulation of Brain Adenylate Cyclase"; Adv. in Cycl. Nucl. Res., 9[1978]85-99.

Brunger, A., Brooks, C. L. III, and Karplus, M.; "Stochastic Boundary Conditions for Molecular Dynamics Simulations of ST2 Water"; Chem. Phys. Let., 105[1984]495-500.

Brunger, A., Brooks, C. L. III, and Karplus, M.; "Active site dynamics of ribonuclease"; Proc. Nat. Acad. Sci., 82[1985]8458-8462.

Bryan, P., Alexander, P, Strausberg, S., Schwartz, F., Lan, W., Gilliland, G., and Gallagher, D. T.; "Energetics of Folding Subtilisin BPN'"; Biochemistry, 31[1992]4937-4945.

Carafoli, E., and Penniston, J. T.; "The Calcium Signal"; Sci. Am., 253[1985]70-78.

Carter, P., Nilsson, B., Burnier, J. P., Burdick, D., and Wells, J. A.; "Engineering Subtilisin BPN' for Site-Specific Proteolysis"; Proteins, 6[1989]240-248.

Carter, P., and Wells, J. A.; "Dissecting the catalytic triad of a serine protease"; Nature, 332[1988]564-568.

Casti, J. L.; "Alternate Realities. Mathematical Models of Nature and Man"; John Wiley and Sons, New York, 1989.

Cornell, W. D., Howard, A. E., and Kollman, P.; "Molecular mechanical potential functions and their application to study molecular systems"; Curr. Opin. Struc. Biol., 1[1991]201-212.

Drenth, J., Hol, W. G. J., Jansonius, J. N., and Koekoek, R.; "Subtilisin Novo: The three dimensional structure and its comparison with subtilisin BPN'"; Eur. J. Biochem., 26[1972]177-181.

Edsell, J. T., and McKensie, H. A.;"Water and Proteins. II. The Location and Dynamics of Water in Protein Systems and Its Relation to Their Stability and Properties"; Adv. Biophys., 16[1983]53-183.

Einspahr, H., and Bugg, C. E.; "Crystal Structures of Calcium Complexes and Implications for Biological Systems"; in "Calcium and its Role in Biology", vol. 17 of "Metal Ions in Biological Systems", ed. Siegel, H., pp 1-49, pub. Marcel Dekker, Inc., New York, 1984.

Estell, D. A., Graycar, T. P., Miller, J. V., Powers, D. B., Burnier, J. P., Ng, P. G., and Wells, J.A.; "Probing Steric and Hydrophobic Effects on Enzyme-Substrate Interactions by Protein Engineering"; Science, 233[1986]659-663.

Factor, A. D., and Mehler, E. L.: "Graphical representation of hydrogen bonding patterns in proteins"; Prot. Eng., 4[1991]421-425.

Fersht, A.; "Enzyme Structure and Mechanism"; W.H. Freeman and Co., New York, 1985.

Franks, F.; "Water: A Comprehensive Treatise"; Vol 44, Plenum, New York, 1979.

Hansen, J-P., and McDonald, I. R.; "Theory of Simple Liquids"; Academic Press, New York, 1986.

Harte, W. E. Jr., Swaminathan, S., Mansuri, M. M., Martin, J. C., Rosenberg, I. E., and Beveridge, D. L.; "Domain communication in the dynamical structure of human immunodeficiency virus 1 protease"; Proc. Nat. Acad. Sci., USA, 87[1990]8864-8868.

Harvey, S. C.; "Treatment of Electrostatic Effects in Macromolecular Modeling"; Proteins, 51[1989]78-92.

Heermann, D. W.; "Computer Simulation Methods in Theoretical Physics"; Springer-Verlag, New York, 1990.

Heiner, A. P., Berendsen, H. J. C., van Gunsteren, W. F.; "MD Simulation of Subtilisin BPN' in a Crystal Environment"; Proteins: Structure, Function and Genetics, 14[1992]451-464.

Herzberg, O., and James, M. N. G.; "Refined Crystal Structure of Troponin C from Turkey Skeletal Muscle at 2.0 Resolution: J. Mol. Biol., 203[1988]761-779.

Hirschfelder, J. D., Curtiss, C. F., and Bird, R. B.; "Molecular Theory of gases and Liquids"; John Wiley and Sons, New York, 1954.

Hirst, D. M.; "The Simulation of Proteins and Macromolecules"; in A Computational Approach to Chemistry, by Hirst, D. M., pub. Blackwell Scientific Pub., Oxford, 1990.

Hori, K., Kushick, J. N., and Weinstein, H.; "Structural and Energetic Parameters of Ca^{2+}-Binding to Peptides and Proteins"; Biopoly., 27[1988]1865-1886.

Howard, A. E., and Kollman, P. A.; "An Analysis of Current Methodologies for Conformational Searching of Complex Molecules"; J. Med. Chem., 31[1988]1669-1675.

Ichiye, T., Karplus, M.; "Collective Motions in Proteins: A Covariance Analysis of Atomic Fluctuations in Molecular Dynamics and Normal Mode Simulations"; Protein, 11[1991]205-217.

Jorgensen, W. L.; "Transferable Intermolecular Potential Functions for Water, Alcohols, and Ethers. Application to Liquid Water"; J. Am. Chem. Soc., 103[1981]335-340.

Karplus, M., and Kushick, J. N.; "Method for Estimating the Configurational Entropy of Macromolecules"; Macromolecules, 14[1981]325-332.

Karplus, M., and McCammon, J. A.; "The Dynamics of Proteins"; Scientific American, 254[1986]42-51.

Karplus, M., and Petsko, G. A.; "Molecular dynamics simulations in biology"; Nature, 347[1990]631-638.

Kartha, G., Varughese, K. I., Aimoto, S.; "Conformation of Cyclo(-L-Pro-Gly-)3 and Its Ca2+ and Mg2+ Complexes"; Proc. Nat. Acad. Sci., 79[1982]4519-4522.

Kottalam, J., and Case, D. A.; "Dynamics of Ligand Escape from the Heme Pocket of Myoglobin"; J. Am. Chem. Soc., 110[1988]7690-7697.

Krauss, M., and Stevens, W. J.; "A Theoretical Model of Metal Binding Sites in Proteins"; in Computer-Assisted Modeling of Receptor-Ligand Interactions: Theoretical Aspects and Applications to Drug Design; pp 95-107, pub. Alan R. Liss, Inc., 1989.

Kraut, J.; "Serine Proteases: Structure and Mechanism of Catalysis"; Ann. Rev. Biochem., 46[1977]331-358.

Kuhn, L. A., Siani, M. A., Pique, M. E., Fisher, C. L., Getzoff, E. D., and Tainer, J. A.;"The Interdependence of Protein Surface Topography and Bound Water Molecules Revealed by Surface Accessibility and Fractal Density Measures"; J. Mol. Biol., 228[1992]13-22.

Leatherbarrow, R. J, and Fersht, A. R.; "Protein Engineering"; Prot. Eng., 1[1986]7-16.

Levitt, M., and Sharon, R.; "Accurate simulation of protein dynamics in solution"; Proc. Nat. Acad. Sci., 85[1982]7557-7561.

Liebman, M. N., Venanzi, C. A., and Weinstein, H.; "Structural Analysis of Carboxypeptidase A and its Complexes with Inhibitors as a Basis for Modeling Enzyme Recognition and Specificity"; Biopoly., 24[1985]1721-1758.

Low, B. W., and Corfield, P. W. R.; "Erabutoxin b. Structure/function relationships following initial protein refinement at 0.140-nm resolution"; Eur. J. Biochem., 161[1986]579-587.

Low, B. W., and Corfield, P. W. R.; "Acetylcholine Receptor: Alpha-Toxin Binding Site - Theoretical and Model Studies"; Asia Pac. J. Pharm., 2[1987]115-127.

Mackay, D. H. J., Cross, A. J., and Hagler, A. T.; "The Role of Energy Minimization in Simulation Strategies of Biomolecular Systems"; in Prediction of Protein Structure and the Principles of Protein Conformation, ed. Fasman, G. D.,; pub. Plenum Press, New York, 1989.

Martin, R. B.; "Bioinorganic Chemistry of Calcium"; in Metal Ions in Biological Systems, ed. Sigel, H., pub. Marcel Dekker, Inc., New York, 1973.

McCammon, J. A., Gelin, B. R., and Karplus, M.; "Dynamics of folded proteins"; Nature, 267[1977]585-590.

McCammon, J. A., and Harvey, S. C.; "Dynamics of Proteins and Nucleic Acids"; Cambridge University Press, New York, 1987.

McQuarrie, D. A.; "Statistical Mechanics"; Harper and Row, New York, 1988.

Moody, P. C. E., and Wilkinson, A. J.; "Protein Engineering"; pub. Oxford University Press, New York, 1990.

Ottesen, M., and Svensen, I.; "The Subtilisins"; Meth. Enzym., 19[1970]199-215.

Oxender, D. L., and Fox, C. F.; "Protein Engineering"; Alan R. Liss, Inc., New York, 1987.

Pantoliano, M. W., Ladner, R. C., Bryan, P. N., Rollence, M. L., Wood, J. F., and Poulos, T. L.; "Protein Engineering of Subtilisin BPN': Enhanced Stabilization through the Introduction of Two Cysteines To Form a Disulfide Bond"; Biochem., 26[1987]2077-2082.

Pantoliano, M. W., Whitlow, M., Wood, J. F., Dodd, S. W., Hardman, K. D., Rollence, M. L., and Bryan, P. N.; "Large Increases in General Stability for Subtilisin BPN' through Incremental Changes in the Free Energy of Unfolding"; Biochem, 28[1989]7205-7213.

Pantoliano, M. W., Whitlow, M., Wood, J. F., Rollence, M. L., Finzel, B. C., Gilliland, G. L., Poulos, T. L., and Bryan, P. N.; "The Engineering of Binding Affinity at Metal Ion Binding Sites for the Stabilization of Proteins: Subtilisin as a Test Case"; Biochem, 27[1988]8311-8317.

Richardson, J. S.; "The anatomy and taxonomy of protein structure"; Adv. Prot. Chem., 34[1981]167-340.

Rigby, M., Smith, E. B., Wakeham, W. A., and Maitland, G. C.; "The Forces Between Molecules"; Oxford University Press, New York, 1986.

Siezen, R. J., de Vos, W. M., Leunissen, A. M., Dijkstra, B. W.; "Homology modelling and protein engineering strategy of subtilases, the family of subtilisin-like serine proteases"; Prot. Eng., 4[1991]719-737.

Smith, E. L., Markland, F. S., and Glazer, A. N.; "Some Structure-Function Relationships in the Subtilisins"; in Structure-Function Relationships of Proteolytic Enzymes, ed. Desnuelle, P., Neurath, H., and Ottesen, M.; pub. Munksgraad, Copenhagen, 1970.

Smith, J. L., Corfield, P. W. R., Hendrickson, W. A., and Low, B. W.; "Refinement at 1.4 Resolution of a Model of Erabutoxin b: Treatment of Ordered Solvent and Discrete Disorder"; Acta Cryst. A, 44[1988]357-368.

Stillinger, F. H., and Rahman, A.; "Improved simulation of liquid water by molecular dynamics"; J. Chem. Phys., 60[1974]1545.

Sussman, F., and Weinstein, H.; "On the Ion Selectivity on Ca-Binding Proteins: Cyclo-(L-Pro-Gly)3 Peptide as a Model"; PNAS 86[1989]7880-7884.

Szebenyi, E. M. E., and Moffat, K.; "Some Thoughts Regarding EF-Hands and the Structure of Calbindins"; in Calcium-binding Proteins in Health and Disease, pp. 323-332; ed. Norman, A. W., et. al., pub. Academic Press, New York, 1987.

Van Gunsteren, W. F.; "The role of computer simulation techniques in protein engineering"; Prot. Eng., 2[1988]5-13.

Van Gunsteren, W. F., and Berendsen, H. J. C.; "Algorithm for macromolecular dynamics andconstraint dynamics"; Mol. Phys., 34[1977]1311.

Verlet, L.; "Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules"; Phys. Rev., 159[1967]98-103.

Wells, J. A., Cunningham, B. C., Graycar, T. P., and Estell, D. A.; "Importance of hydrogen-bond formation in stabilizing the transition state of subtilisin"; Phil. Trans. R. Soc. Lond. A, 317[1986]415-423.

Wells, J. A., Cunningham, B. C., Graycar, T. P., Estell, D. A., and Carter, P.; "On the Evolution of Specificity and Catalysis in Subtilisin"; Cold Spring Harbor Symposia on Quantitative Biology, 52[1987a]647-652.

Wells, J. A., and Estell, D.; "Subtilisin: An Enzyme Designed to be Engineered"; TIBS, 13[1988a]291-297.

Wells, J. A., Powers, D. B., Bott, R. R., Graycar, T. P., and Estell, D. A.; "Designing substrate specificity by protein engineering of electrostatic interactions"; Proc. Nat. Acad. Sci., USA, 84[1987b]1219-1223.

Wells, J. A., Powers, D. B., Bott, R. R., Katz, B. A., Ultsch, M. H., Kossiakoff, A. A., Power, S. D., Adams, R. M., Heyneker, H. H., Cunningham, B. C., Miller, J. V., Graycar, T. P., and

Estell, D. A.; "Protein Engineering of Subtilisin"; in Protein Engineering, ed. Oxender, D. L., and Fox, C. F.; pub. Alan R. Liss, Inc., New York, 1987c.

Wells, J. A., Vasser, M., and Powers, D. B.; "Cassette mutagenesis: an efficient method for generation of multiple mutations at defined sites"; Gene, 34[1985]315-323.

Wells, J. A., Cunningham, B. C., Graycar, T. P., and Estell, D. A.; "Recruitment of substrate-specificity properties from one enzyme into a related one by protein engineering"; PNAS, 84[1987]5167-5171.

Williams, R. J. P.; "The physics and chemistry of the calcium-binding proteins"; pp145-161, in "Calcium and the Cell", John Wiley and Sons, NY, 1986

Willis, B. T. M., and Pryor, A. W.; "The Atomic Temperature Factor (Debye-Waller Factor)"; in Thermal Vibrations in Chemistry, pp 81-141; pub. Cambridge U. Press, Cambridge, 1975.

Wright, C. S., Alden, R. A., and Kraut, J.; "Structure of subtilisin at 2.5Å resolution"; Nature, 221[1969]235-242.